

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 10, 2005, 06:08:13 ; Search time 370 Seconds  
(without alignments)  
4984.946 Million cell updates/sec

Title: US-10-647-089-93  
Perfect score: 321  
Sequence: 1 ttgaccacagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	321	100.0	321	9	US-09-894-844-93
2	321	100.0	321	16	US-10-388-902-93
3	321	100.0	321	16	US-10-647-089-93
c	44.8	14.0	1599	15	US-10-156-761-2997
	44.8	14.0	9025608	15	US-10-156-761-1
	44.8	13.7	909	18	US-10-425-115-73640
	43.8	13.6	3480	15	US-10-369-493-39466
7	43.8	13.6	3480	15	US-10-369-493-39849
8	43.8	13.6	3501	15	US-10-369-493-39098
9	43.8	13.6	3501	15	US-10-369-493-39098
10	43	13.4	1193	16	US-10-425-114-24011
11	43	13.4	1308	18	US-10-425-115-49576
12	42.8	13.3	4200	9	US-09-815-242-4124

C	13	42.8	13.3	4200	16	US-10-282-122A-7392	Sequence 7392, Ap
	14	42.8	13.3	10232	16	US-10-282-122A-25487	Sequence 25487, A
C	15	42	13.1	1959	15	US-10-156-761-4513	Sequence 4513, Ap
	16	42	13.1	9025608	15	US-10-156-761-1	Sequence 1, Appli
	17	41.6	13.0	996	16	US-10-282-122A-11268	Sequence 11268, A
C	18	41.2	12.8	936	17	US-10-437-963-76527	Sequence 76527, A
	19	41	12.8	400	9	US-09-867-701-4968	Sequence 4968, Ap
	20	40.8	12.7	1232	17	US-10-437-963-29422	Sequence 29422, A
	21	40.6	12.6	2580	15	US-10-156-761-3895	Sequence 3895, Ap
C	22	40.6	12.6	3044	17	US-10-437-963-79637	Sequence 79637, A
	23	40.4	12.6	951	16	US-10-282-122A-26156	Sequence 26156, A
	24	40.4	12.6	954	9	US-09-712-363-39	Sequence 39, Appl
	25	40.4	12.6	954	16	US-10-282-122A-28300	Sequence 28300, A
	26	40.4	12.6	45191	15	US-10-080-170-649	Sequence 649, App
	27	40.4	12.6	45191	17	US-10-080-170-649	Sequence 649, App
	28	40.4	12.6	45191	18	US-10-468-356-649	Sequence 649, App
	29	40.2	12.5	307	18	US-10-425-115-131105	Sequence 131105, A
C	30	40.2	12.5	1803	15	US-10-156-761-4657	Sequence 4657, Ap
	31	40.2	12.5	1824	17	US-10-767-701-13070	Sequence 13070, A
C	32	39.8	12.4	1300	17	US-10-437-963-22867	Sequence 22867, A
	33	39.6	12.3	1318	16	US-10-425-114-14884	Sequence 14884, A
C	34	39.6	12.3	1446	15	US-10-156-761-4324	Sequence 4324, Ap
	35	39.6	12.3	1460	16	US-10-425-114-14893	Sequence 14893, A
C	36	39.6	12.3	1512	13	US-10-124-800-31	Sequence 31, Appl
	37	39.6	12.3	1512	18	US-10-810-352-31	Sequence 31, Appl
C	38	39.6	12.3	2073	16	US-10-425-114-29125	Sequence 29125, A
	39	39.6	12.3	4509	13	US-10-124-800-5	Sequence 5, Appli
C	40	39.6	12.3	4509	18	US-10-810-352-5	Sequence 5, Appli
	41	39.6	12.3	4512	13	US-10-124-800-27	Sequence 27, Appl
C	42	39.6	12.3	4512	18	US-10-810-352-27	Sequence 27, Appl
	43	39.6	12.3	4767	15	US-10-331-061-76	Sequence 76, Appl
C	44	39.4	12.3	1488	9	US-09-815-242-7862	Sequence 7862, Ap
	45	39.4	12.3	1488	16	US-10-282-122A-30375	Sequence 30375, A

ALIGNMENTS

RESULT 1  
US-09-894-844-93  
; Sequence 93, Application US/09894844  
; Patent No. US20020176873A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of the M. Tuberculosis Complex  
; FILE REFERENCE: STAN102CON  
; CURRENT APPLICATION NUMBER: US/09/894,844  
; CURRENT FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 93  
; LENGTH: 321  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
US-09-894-844-93

Query Match 100.0%; Score 321; DB 9; Length 321;  
Best Local Similarity 100.0%; Pred. No. 9.3e-90;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTGACCCACAGCGCAGCTAAACGCCAGCGAGGCTCGCCGAGGCTCAAGCCCGCGT 60  
DB 1 TTGACCCACAGCGCAGCTAAACGCCAGCGAGGCTCGCCGAGGCTCAAGCCCGCGT 60  
QY 61 CGGAATCGGTTGGCGCGCAACATGGTTGGCGGCGGACGTTTCGTCGCGGAGCAGCGC 120

Db 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCGTCGCGGAGCAGCG 120  
QY 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGAGCGGTACGCGGAGATGGTGGCG 180  
Db 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGAGCGGTACGCGGAGATGGTGGCG 180  
QY 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
Db 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
QY 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
Db 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
QY 301 GACGAACCCGACGCAACACAG 321  
Db 301 GACGAACCCGACGCAACACAG 321

## RESULT 2

US-10-388-902-93  
; Sequence 93, Application US/10388902  
; Publication No. US20040018574A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; FILE REFERENCE: STAN102CON  
; CURRENT APPLICATION NUMBER: US/10/388,902  
; PRIOR FILING DATE: 2003-03-14  
; PRIOR APPLICATION NUMBER: US/09/894,844  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 93  
; LENGTH: 321  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
US-10-388-902-93

Query Match 100.0%; Score 321; DB 16; Length 321;  
Best Local Similarity 100.0%; Pred. No. 9.3e-90;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTGACCCACAGCGCACTAAACGCCAGCGAGCATCGCGCAGGGCTCAACGCCCGCGT 60  
Db 1 TTGACCCACAGCGCACTAAACGCCAGCGAGCATCGCGCAGGGCTCAACGCCCGCGT 60  
QY 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCGTCGCGGAGCAGCGC 120  
Db 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCGTCGCGGAGCAGCGC 120  
QY 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGCGCGAGCTTCGTCGCGGATGGTGGCG 180  
Db 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGCGCGAGCTTCGTCGCGGATGGTGGCG 180  
QY 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
Db 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
QY 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
Db 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
QY 301 GACGAACCCGACGCAACACAG 321

Db 301 GACGAACCCGACGCAACACAG 321  
RESULT 3  
US-10-647-089-93  
; Sequence 93, Application US/10647089  
; Publication No. US20040063923A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; FILE REFERENCE: the M. Tuberculosis Complex  
; CURRENT APPLICATION NUMBER: US/10/647,089  
; PRIOR FILING DATE: 2003-08-21  
; PRIOR APPLICATION NUMBER: US/09/894,844  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 93  
; LENGTH: 321  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
US-10-647-089-93

Query Match 100.0%; Score 321; DB 16; Length 321;  
Best Local Similarity 100.0%; Pred. No. 9.3e-90;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTGACCCACAGCGCACTAAACGCCAGCGAGCATCGCGCAGGGCTCAACGCCCGCGT 60  
Db 1 TTGACCCACAGCGCACTAAACGCCAGCGAGCATCGCGCAGGGCTCAACGCCCGCGT 60  
QY 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCGTCGCGGAGCAGCGC 120  
Db 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCGTCGCGGAGCAGCGC 120  
QY 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGCGCGAGCTTCGTCGCGGATGGTGGCG 180  
Db 121 CGCGCCCAACGGCAGCGGACCTCAGGCTATCCGCGCGCGAGCTTCGTCGCGGATGGTGGCG 180  
QY 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
Db 181 ACATCACAGAAATCGAGCGAGCACAGCCGAACTGGCGTGTGTGATGCAATCTCGAC 240  
QY 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
Db 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGATCACTTCCCC 300  
QY 301 GACGAACCCGACGCAACACAG 321  
Db 301 GACGAACCCGACGCAACACAG 321

## RESULT 4

US-10-156-761-2997/c  
; Sequence 2997, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA

Query Match 14.0%; Score 44.8; DB 15; Length 9025608;  
Best Local Similarity 51.5%; Pred. No. 0.00078;  
Matches 103; Conservative 0; Mismatches 97; Indels 0; G

RESULT 7  
US-10-369-493-39466  
; Sequence 39466, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkie, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 39466  
; LENGTH: 3480  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-39466

Query Match 13.6%; Score 43.8; DB 15; Length 3480;  
Best Local Similarity 55.6%; Pred. No. 0.0011;  
Matches 84; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 111 CGAGCAGCGCGCCCAACGGCAGCGACCTCGAGGCTATCGCGAGCGTACGCCGA 170  
DB 924 CCAGCACCAGCGGAATCTCGCATCGCTGCAAGCGCGGCGACGAAGCGGAGGCCA 983  
QY 171 GATGGTGGCGACATCACAGAAATCGACGACGACACAGCCGAACTGGCGCTGTGTGCGAT 230  
DB 984 GCTGCAGGAGCTACCCAGCACATCAGCGGCGACTCGCCCAAGCTGGCGGTGTGCGCGA 1043  
QY 231 GCATCTCGACGATGAGCAGCGCGCGCTTGAG 261  
DB 1044 GCGGTGGCGGATGCCGAACCGAGCTGGAG 1074

## RESULT 8

US-10-369-493-39849  
; Sequence 39849, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 39849  
; LENGTH: 3480  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-39849

Query Match 13.6%; Score 43.8; DB 15; Length 3480;  
Best Local Similarity 55.6%; Pred. No. 0.0011;  
Matches 84; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 111 CGAGCAGCGCGCCCAACGGCAGCGACCTCGAGGCTATCGCGAGCGTACGCCGA 170  
DB 924 CCAGCACCAGCGGAATCTCGCATCGCTGCAAGCGCGGCGACGAAGCGGAGGCCA 983  
QY 171 GATGGTGGCGACATCACAGAAATCGACGACGACACAGCCGAACTGGCGCTGTGTGCGAT 230  
DB 984 GCTGCAGGAGCTACCCAGCACATCAGCGGCGACTCGCCCAAGCTGGCGGTGTGCGCGA 1043  
QY 231 GCATCTCGACGATGAGCAGCGCGCGCTTGAG 261  
DB 1044 GCGGTGGCGGATGCCGAACCGAGCTGGAG 1074

## RESULT 9

US-10-369-493-39098  
; Sequence 39098, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 39098  
; LENGTH: 3501  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-39098

Query Match 13.6%; Score 43.8; DB 15; Length 3501;  
Best Local Similarity 55.6%; Pred. No. 0.0011;  
Matches 84; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 111 CGAGCAGCGCGCCCAACGGCAGCGACCTCGAGGCTATCGCGAGCGTACGCCGA 170  
DB 924 CCAGCACCAGCGGAATCTCGCATCGCTGCAAGCGCGGCGACGAAGCGGAGGCCA 983  
QY 171 GATGGTGGCGACATCACAGAAATCGACGACGACACAGCCGAACTGGCGCTGTGTGCGAT 230  
DB 984 GCTGCAGGAGCTACCCAGCACATCAGCGGCGACTCGCCCAAGCTGGCGGTGTGCGCGA 1043  
QY 231 GCATCTCGACGATGAGCAGCGCGCGCTTGAG 261  
DB 1044 GCGGTGGCGGATGCCGAACCGAGCTGGAG 1074

## RESULT 10

US-10-425-114-24011  
; Sequence 24011, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114  
; PRIOR FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 24011  
; LENGTH: 1193  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3606-060-D3\_FLI  
US-10-425-114-24011

Query Match 13.4%; Score 43; DB 16; Length 1193;  
Best Local Similarity 50.2%; Pred. No. 0.0018;  
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 107 CCGCCGAGCAGCGCGCGCCCAACGGCAGCGGACCTCGAGGCTATCGCGAGCGTACG 166  
DB 331 CCGCGGAGATCAACCTGTCTCCCTGCCCGGGGACGCTTCGACCGCTGGCCCGCAAGG 390  
QY 167 CCGAGATGTTGGCGACATCACAGAAATCGACGACGACACAGCCGAACTGGCGCTGTGCT 226  
DB 391 CCGCGCTGAGGCGCGCTGTGCGCGGCGTGTCTCTGCGCGACGCGCTGGCGCTAGCGG 450  
QY 227 CGATCATCTTCGACGATGAGCAGCGCGCGCTTGAAGCGGGGATGAAGCTCGGCTCGCATC 286

Db 451 CGCGGACCTCGTGGCCCTCGGGCGCCCGGTTCCCGTGGCGTGGCGCGCGCG 510  
Qy 287 CGTATCACTCCCGACGAAACCCGACACAA 317  
Db 511 ACTCCCGCGCTCCGACGCGCGCGACGTCGA 541

## RESULT 11

US-10-425-115-49576  
; Sequence 49576, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 49576  
; LENGTH: 1308  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MXT4577\_145214C.1  
US-10-425-115-49576

Query Match 13.4%; Score 43; DB 18; Length 1308;  
Best Local Similarity 50.2%; Pred. No. 0.0018;  
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
Qy 107 CGCGGACGCGCGCGCGCCCAACGGGACGGGACCTCGAGGCTATCCGCCGAGGTCAG 166  
Db 331 CGCGGAGATCAACCTGCTCCCTCGCCCGGGGACGCTTCGACGCGCTGGCCCGGCCAAGG 390  
Qy 167 CCGAGATGTCGGGACATCACAGAAATCGACGACACAGCGAACTGGCGCTGTGT 226  
Db 391 CGCGCTGGAGCGCGGTGCGCGCGCTGCTCTCGCCGACGCGCTGGCGCTAGCGG 450  
Qy 227 CGATGATCTCGACGATGAGCAGCGCGCTTGAGCGCGGGATGAAGCTCGGCTGGCATC 286  
Db 451 CGCGGACCTCGTGGCGCCCTTGGCGGCGCCCGGTTCCCGTTCGCGTGGCGCGCGCG 510  
Qy 287 CGTATCACTCCCGACGAAACCCGACACAA 317  
Db 511 ACTCCCGCGCTCCGACGCGCGCGACGTCGA 541

## RESULT 12

US-09-815-242-4124/c  
; Sequence 4124, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 4124  
; LENGTH: 4200  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-4124

Query Match 13.3%; Score 42.8; DB 9; Length 4200;  
Best Local Similarity 50.5%; Pred. No. 0.0022;  
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
Qy 48 CAACGCCCCGCTCGGAATCGGTTGGCGGCAACATGGTTGGCGCGCGCGCTCCGTC 107  
Db 1519 CCACGCTCTCGGACGGAACGATGATCGGCTCGCGCTTGGCGGCGGACAGGATGTTGTTGG 1460  
Qy 108 CGCGGACGCGCGCGCCCAACGGGACGGGACCTCGAGGCTATCCGCCGAGCGTACGC 167  
Db 1459 TGGACATCATCAGCGCGCGCGCTTCCAGCTGGGCTCGAGGTCAGCGGAACGTTGACGG 1400  
Qy 168 CGAGATGTCGGGACATCACAGAAATCGACGACACAGCGAACTGGCGCTGTGTC 227  
Db 1399 CCATCTGTCACCGTCGAAGTCGCGCTTGATCGGCGGCGACAGCGGTCGAGCTGGA 1340  
Qy 228 GATCATCTCGACCATGAGCGGCC 253  
Db 1339 TCGCCTTTCCTTCGATGAGGACCGCG 1314

## RESULT 13

US-10-282-122A-7392/c  
; Sequence 7392, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7392
; LENGTH: 4200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-7392

Query Match      13.3%; Score 42.8; DB 16; Length 4200;
Best Local Similarity 50.5%; Pred. No. 0.0022;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 48 CAACGCCCGCGTCCGATCGGTTGGCGGCAACATGTTGGCGGCGACGTTCCGTC 107
Db 1519 CCACGTCCTCGACGGAACATGATCGGCTCGCGGCGGCGACAGGATGTTG 1460

QY 108 CGCGGACGAGCGCGCGCCCAACGCGAGCGGACCTCGAGGCTATCCGCCGAGCGTACGC 167
Db 1459 TGGACATCATCAGCGCGCGCTTCCAGCTGGCGCTCGAGGTCAGGAGNACGTGACGG 1400

QY 168 CGAGATGTTGGGACATCAACAGAAATCGACGACGACACAGCGAACTGGCGCTGTGTC 227
Db 1399 CCATCTGTCACCGTCGAAGTCGCGGTTGTACGCGGCGGACAGCCAGCGGTGACGCTGA 1340

QY 228 GATGCATCTCGACGATGAGCAGCGCC 253
Db 1339 TCGCCTTGCCTTCGATGAGGACCGGC 1314

RESULT 14
US-10-282-122A-25487
; Sequence 25487, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 25487
; LENGTH: 10232
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-25487

Query Match      13.3%; Score 42.8; DB 16; Length 10232;
Best Local Similarity 47.7%; Pred. No. 0.0023;
Matches 125; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 20 AAGCGCAGCCAGCCATCCGCGCAGGCTCAACGCGCCCGGTGGAATCGCGTTGGCGGC 79
Db 1271 AACTGCTGACCGCGGACGAGCCCGCGCTGATCGACCGGTGGAGCAACCGTCCGCGCTGA 1330

QY 80 AACATGTGTTGGCGCGCCGACGTTCCGTCGCGGAGCAGCGCGCCCAACGCGCAGCGC 139
Db 1331 CCGAGCCCGCACCCCGCGGTGTCGATCCCGAGCCTTCGCCGACACGTCGAGCGCA 1390

QY 140 ACCTGAGGCTATCCGCGAGGCTAGCGGAGATGTTGGGACATCACACGAAATCGAGC 199
Db 1391 CCCCCGACGCGGTGGCGGTGACGTTCCGGGCGACCTCGCTGACCTACGCCAGCTCGAGC 1450

QY 200 AGACACACAGCCAACTGGCGCTGTTGTGATGTCATCTCGACGATGAGCAGCGCGCTTG 259
Db 1451 AGCGTCCACCGGCTGGGCCATCTGTCGCGCACACGCGGTGGCGCGCGGCGACTGCG 1510

QY 260 AGCGCGGATGAAGTCCGCGTG 281
Db 1511 TGGCGGTGATGTTCCCGCGCTG 1532

RESULT 15
US-10-156-761-4513
; Sequence 4513, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4513
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1959)
US-10-156-761-4513

Query Match      13.1%; Score 42; DB 15; Length 1959;
Best Local Similarity 50.5%; Pred. No. 0.0038;
Matches 102; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

Qy	20	AACGCCAGCCAGCCATCGCCGAGGGCTCAACGCCCCCGCGTCGGAATCGCGTTGGCGGC	79
Db	365	AGCAGGACACCTCAACCGTCAACGACCTGGGGGACGCCCGTCGGCGCGCTTCGGGGGCC	424
Qy	80	AACATGGTTGGCCGCCGACGTTCCGTCGGCCGAGCAGCGCGGCCCAACGGCAGCGCG	139
Db	425	GGCAGGAGCCCGACCCCGACCTGCGCGCCCTTCCCGCGCCACGCGCCCGCCCGCCG	484
Qy	140	ACCTCGAGGCTATCGCCGAGCGTACGCCGAGATGGTGGCGACATCACACGAAATCGACG	199
Db	485	TCGAGGACGACCCCGAGCATGACACCCGCGCATCTGGCGCCCGGACCCAGTTCGACC	544
Qy	200	ACGACACAGCCGAACTGGCGCT	221
Db	545	CGTGGTTCGCGGACCTGGTGGT	566

Search completed: February 10, 2005, 07:36:14  
Job time : 390 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 05:25:02 ; Search time 75 Seconds  
(without alignments)  
3042.176 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321

Sequence: 1 ttgaccacagcgcaactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	321	100.0	321	4	US-09-894-844-93
2	321	100.0	4403765	3	US-09-103-840A-2
3	321	100.0	4411529	3	US-09-103-840A-1
4	42.8	13.3	4212	4	US-09-252-991A-4748
5	42.8	13.3	4242	4	US-09-252-991A-4697
6	40.4	12.6	4403765	3	US-09-103-840A-2
7	40.4	12.6	4411529	3	US-09-103-840A-1
8	40.2	12.5	732	4	US-09-252-991A-6298
9	40.2	12.5	945	4	US-09-252-991A-6167
10	40.2	12.5	1602	4	US-09-252-991A-6382
11	39.8	12.4	594	4	US-09-252-991A-12043
12	39.8	12.4	966	4	US-09-252-991A-11896
13	39.8	12.4	1179	4	US-09-252-991A-12017
14	39.6	12.3	4767	4	US-09-231-899-76
15	39.4	12.3	1470	4	US-09-252-991A-6325
16	39.4	12.3	1473	4	US-09-252-991A-6144
17	39.4	12.3	1512	4	US-09-252-991A-6219
18	39.2	12.2	1320	4	US-09-724-797-85
19	38.8	12.1	408	4	US-09-252-991A-655
20	38.8	12.1	801	4	US-09-252-991A-698
21	38.8	12.1	1356	4	US-09-252-991A-630
22	38.8	12.1	2955	4	US-09-252-991A-684
23	38.8	12.1	3150	4	US-09-252-991A-727
24	38.4	12.0	1371	4	US-09-252-991A-13954
25	38.4	12.0	2181	4	US-09-252-991A-13742
26	38.4	12.0	2277	4	US-09-252-991A-13403
27	38.4	12.0	2549	4	US-09-638-203-1

C 28	38.4	12.0	2570	4	US-10-140-002-539	Sequence 539, Appl
29	38.2	11.9	888	4	US-09-252-991A-16512	Sequence 16512, A
30	38.2	11.9	1512	4	US-09-252-991A-16092	Sequence 16092, A
31	38.2	11.9	2400	4	US-09-252-991A-15991	Sequence 15991, A
32	37.8	11.8	4257	2	US-08-690-473-1	Sequence 1, Appli
33	37.8	11.8	4257	3	US-09-259-821A-1	Sequence 1, Appli
34	37.8	11.8	4257	3	US-08-843-659-1	Sequence 1, Appli
35	37.8	11.8	4257	4	US-08-825-288A-1	Sequence 11, Appl
36	37.8	11.8	12001	1	US-08-458-568A-11	Sequence 943, App
37	37.2	11.6	1488	4	US-09-489-039A-943	Sequence 896, App
38	37	11.5	1470	4	US-09-252-991A-896	Sequence 1137, Ap
39	37	11.5	1662	4	US-09-252-991A-1137	Sequence 23, Appl
40	37	11.5	1693	3	US-09-320-878-23	Sequence 19, Appl
41	37	11.5	1693	4	US-09-141-908-19	Sequence 23, Appl
42	37	11.5	1693	4	US-09-657-440-23	Sequence 945, App
43	37	11.5	1839	4	US-09-252-991A-945	Sequence 7, Appli
44	37	11.5	44377	2	US-08-804-227C-7	Sequence 1, Appli
45	37	11.5	44377	2	US-08-804-198-1	

## ALIGNMENTS

RESULT 1  
US-09-894-844-93  
; Sequence 93, Application US/09894844  
; Patent No. 6686166  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; FILE REFERENCE: STAN102CON  
; CURRENT APPLICATION NUMBER: US/09/894,844  
; PRIORITY FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 93  
; LENGTH: 321  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
US-09-894-844-93

Query Match	100.0%	Score 321;	DB 4;	Length 321;
Best Local Similarity	100.0%	Pred. No. 3.6e-75;		
Matches 321;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TTGACCCCAACGCGCAGCTAAACGCGAGCGAGCGCTACGCGCGGCTCAACGCCCGCGT	60	
Db	1	TTGACCCCAACGCGCAGCTAAACGCGAGCGAGCGCTACGCGCGGCTCAACGCCCGCGT	60	
Qy	61	CGGAATCGCGTTGGCGGCAACATGTTGGCGCGGCGAGCTTCGCTCCGCGAGCAGCGC	120	
Db	61	CGGAATCGCGTTGGCGGCAACATGTTGGCGCGGCGAGCTTCGCTCCGCGAGCAGCGC	120	
Qy	121	CGCGCCCAACGCGCAGCGAGCTCGAGGCTATCGCGCGGAGCGTACGCCGAGATGGTGGC	180	
Db	121	CGCGCCCAACGCGCAGCGAGCTCGAGGCTATCGCGCGGAGCGTACGCCGAGATGGTGGC	180	
Qy	181	ACATCACAGAAATCGACGACGACACAGCGAGCTGGCGCTGTTGTCATGATCTCGAC	240	
Db	181	ACATCACAGAAATCGACGACGACACAGCGAGCTGGCGCTGTTGTCATGATCTCGAC	240	
Qy	241	GATGAGCAGCGCGCGCTTTGAGCGGGATGAAGCTCGGCTCGCATCCGTCATCATTCCCC	300	
Db	241	GATGAGCAGCGCGCGCTTTGAGCGGGATGAAGCTCGGCTCGCATCCGTCATCATTCCCC	300	



```

Qy 168 CGAGATGGTGGCGACATCACACGAATCGACGACGACACAGCCGAACCTGGCGCTGTTGTC 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1411 CCATCTGGTCCACCGTCCGAAGTCGCGCGTTGTACGCGCGCAGACACAGCGGTGCAGCTGGA 1352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 228 GATGCATCTCGACGATGACGACGAGCGCC 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1351 TCGCCTTGCCTTCGATGAGGACCGGC 1326
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-09-252-991A-4697
; Sequence 4697, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4697
; LENGTH: 4242
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4697

```

Query Match	13.3%	Score 42.8;	DB 4;	Length 4242;
Best Local Similarity	50.5%;	Pred. No. 0.011;		
Matches 104;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;
Qy	48	CAACGCCCGCGTTCGGAATCGGTTGGCGCGCAACATGGTTGGCGCGCCGACGTTCCGTC	107	
Db	2685	CCAGTCTCTGGACGGAAACGATGATCGGTCGCCGTTGGCGGGCGACAGATGTTGTTGG	2744	
Qy	108	CGCCGACGAGCGCGCGCCCAACGGCAGCGCGACCTCGAGGCTATTCGCGCAGGCGTACGC	167	
Db	2745	TGGACATCATCAGCGCGCGGCTTCCAGCTGGGCTTCGAGGGTCAGCGAACCTGGACCG	2804	
Qy	168	CGAGATGTTGGCGGACATCACACGAAATCGACGACACACGCGAAGCTGGCGGCTGTTGTC	227	
Db	2805	CCATCTGGTCACCGTTCGAAGTCGCGGTTGTACGGCGGCGCAACACGCGGTTGAGCTGGA	2864	
Qy	228	GATGCATCTCGACGATGACGAGGCC	253	
Db	2865	TCGCGCTTGGCTTCGATGAGGACCGGC	2890	

```

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent NO. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:

```

[illegible]

RESULT 7  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R. C.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match	12.6%;	Score 40.4;	DB 3;	Length 4411529;
Best Local Similarity	50.4%;	Pred. No. 0.53;		
Matches 129;	Conservative	0;	Mismatches 121;	Indels 6; Gaps 1;
Qy	72	TGGCGGCGCAACATGGTTGGCGGCGGACAGCTTCGTCGCGCGGAGCAGCGCGCGGCCAACG	131	
Db	1129425	TCGGCTTGCACACAGACCGTGGCGGAGCACTCGCACGCGAGGTTACCGACTAACGGTGT	1129484	
Qy	132	GCAGCGCGACCTCGAGGCTATCGCGGACGGTA-----CGCGGAGATGGTGGGACATC	185	
Db	1129485	CAATCGCGACGTCTCGCCCTTCGCGCGGAGGATCTAGCCGCGGCGCCAGCCGGTGT	1129544	
Qy	186	ACACGAAATCGACGACGACACAGCCGAACTGGCGCTGTTGTGATGCATCTCGACGATGA	245	
Db	1129545	TGCCATCTGCCGTACAA CGTAGCGGTACCGGGTGTGTCATCTGCTTGTGAGTTCCC	1129604	
Qy	246	GCAGCGCGGCTTGAGCGCGGGAGTGAAGCTCGGCTGGCATTCGTTATCACTTCCCGACGA	305	
Db	1129605	GTCCGATCCGTGTCGACCGTGATGTGAGGCGGAGGTCGCCAACGGCTCGCCCGCA	1129664	

QY 306 ACCGACAGCAACAG 321  
 Db 1129665 GCCGGGACAAAG 1129680

RESULT 8  
 US-09-252-991A-6298/c  
 ; Sequence 6298, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6298  
 ; LENGTH: 732  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6298

Query Match 12.5%; Score 40.2; DB 4; Length 732;  
 Best Local Similarity 49.3%; Pred. No. 0.1;  
 Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGACGAGCGCCGCCAACGCGACGCGGACCTCGAGGC 149  
 Db 692 GCGATCGCGCTTGGGACAGCGCCACACAGCGCTCGCCGCCAGCGCCACCTCGA 633

QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGGACATCACAGAAATCGACGACACAGC 209  
 Db 632 TGCGGACCTGGCTTCCGGCGGCGGAGCGCCGCCAGCCACATCGATCGCGCTTCGG 573

QY 210 CGAATCGGCGCTTGTTCGATGATCTCGACGATGAGCAGCGCGCTTGGAGCGGGGAT 269  
 Db 572 CGATCTGCGCCCATTTGCGCGCGGAGCGCTTTCGAGCAGCTCCAGTTGCACCTCGGAT 513

QY 270 GAAGCTCGGCTGGCATCCGTATCACTTCCCGA 302  
 Db 512 GCGCTCGGGAACAGTCGATCGCCTCGTGA 480

RESULT 9  
 US-09-252-991A-6167  
 ; Sequence 6167, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6167  
 ; LENGTH: 945  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6167

Query Match 12.5%; Score 40.2; DB 4; Length 945;  
 Best Local Similarity 49.3%; Pred. No. 0.11;  
 Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGACGAGCGCCGCCAACGCGACGCGGACCTCGAGGC 149  
 Db 420 GCGATCGCGCTTGGGACAGCGCCACACAGCGGCTCGCCGCCAGCGCCACCTCGA 479

QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGGACATCACAGAAATCGACGACACAGC 209  
 Db 480 TGCGGACCTGGCTTCCGGCGGCGGAGCGCCGCCAGCCACATCGATCGCGCTTCGG 539

QY 210 CGAATCGGCGCTTGTTCGATGATCTCGACGATGAGCAGCGCGCTTGGAGCGGGGAT 269  
 Db 540 CGATCTGCGCCCATTTGCGCGCGGAGCGCTTTCGAGCAGCTCCAGTTGCACCTCCGAT 599

QY 270 GAAGCTCGGCTGGCATCCGTATCACTTCCCGA 302  
 Db 600 GCGCTCGGGAACAGTCGATCGCCTCGTGA 632

RESULT 10  
 US-09-252-991A-6382/c  
 ; Sequence 6382, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6382  
 ; LENGTH: 1602  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6382

Query Match 12.5%; Score 40.2; DB 4; Length 1602;  
 Best Local Similarity 49.3%; Pred. No. 0.12;  
 Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGACGAGCGCCGCCAACGCGAGCGGACCTCGAGGC 149  
 Db 1222 GCGATCGCGCTTGGGACAGCGCCACACAGCGGCTCGCCGCCAGCGCCACCTCGA 1163

QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGGACATCACAGAAATCGACGACACAGC 209  
 Db 1162 TGCGGACCTGGCTTCCGGCGGCGGAGCGCCGCCAGCCACATCGATCGCGCTTCGG 1103

QY 210 CGAATCGGCGCTTGTTCGATGATCTCGACGATGAGCAGCGCGCTTGGAGCGGGGAT 269  
 Db 1102 CGATCTGCGCCCATTTGCGCGCGGAGCGCTTTCGAGCAGCTCCAGTTGCACCTCCGAT 1043

QY 270 GAAGCTCGGCTGGCATCCGTATCACTTCCCGA 302  
 Db 1042 GCGCTCGGGAACAGTCGATCGCCTCGTGA 1010

RESULT 11  
 US-09-252-991A-12043  
 ; Sequence 12043, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6167  
 ; LENGTH: 945  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6167

Query Match 12.5%; Score 40.2; DB 4; Length 945;  
 Best Local Similarity 49.3%; Pred. No. 0.11;  
 Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12043
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12043

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 594;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCGCCGAGGCTCAACGCCGCCGTCGGAATCGCGTTGGGGGCAACATG 85
DB 265 AGGAAGCCTTCGCTCGGGCCGCTGGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 324
QY 86 GTTGGCCGCGGACGTTTCGTCGCGGAGCAGCGCGGCCCAACAGCGGAGCGGACCTCG 145
DB 325 ACTTCCGAGCTGTCGAGGCCGAGCGCGATGAAGGCGTGGCGAAGCGGCGCTGGTGC 384
QY 146 AGGCTATCGCGGAGGCTACGCCGAGATGGTGGC 180
DB 385 AGGGTTTCTGCGCCTGTTCTCTCGCTGAGGTGGC 419

RESULT 12
US-09-252-991A-11896/c
; Sequence 11896, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11896
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11896

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 966;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCGCCGAGGCTCAACGCCGCCGTCGGAATCGCGTTGGGGGCAACATG 85
DB 584 AGGAAGCCTTCGCTCGGGCCGCTGGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 525
QY 86 GTTGGCCGCGGACGTTTCGTCGCGGAGCAGCGCGGCCCAACAGCGGAGCGGACCTCG 145
DB 524 ACTTCCGAGCTGTCGAGGCCGAGCGCGATGAAGGCGTGGCGAAGCGGCGCTGGTGC 465
QY 146 AGGCTATCGCGGAGGCTACGCCGAGATGGTGGC 180
DB 464 AGGGTTTCTGCGCCTGTTCTCTCGCTGAGGTGGC 430

RESULT 13
US-09-252-991A-12017
; Sequence 12017, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12017
; LENGTH: 1179
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12017

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 1179;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCGCCGAGGCTCAACGCCGCCGTCGGAATCGCGTTGGGGGCAACATG 85
DB 482 AGGAAGCCTTCGCTCGGGCCGCTGGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 541
QY 86 GTTGGCCGCGGACGTTTCGTCGCGGAGCAGCGCGGCCCAACAGCGGAGCGGACCTCG 145
DB 542 ACTTCCGAGCTGTCGAGGCCGAGCGCGATGAAGGCGTGGCGAAGCGGCGCTGGTGC 601
QY 146 AGGCTATCGCGGAGGCTACGCCGAGATGGTGGC 180
DB 602 AGGGTTTCTGCGCCTGTTCTCTCGCTGAGGTGGC 636

RESULT 14
US-09-231-899-76/c
; Sequence 76, Application US/09231899
; Patent No. 6566583
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael
; APPLICANT: Metz, James G
; APPLICANT: Facciotti, Daniel
; TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
; FILE REFERENCE: CGNE.131.0205
; CURRENT APPLICATION NUMBER: US/09/231,899
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; EARLIER APPLICATION NUMBER: 09/090,793
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Schizochytrium aggregatum
US-09-231-899-76

Query Match
Best Local Similarity 12.3%; Score 39.6; DB 4; Length 4767;
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 70 GTTGGGGGCAACATGGTTGGCCGCCGACGTTCCGTCGCCGAGCAGCGCGGCCCAA 129
DB 4026 GGTCCGAGCTGCTTGGCGACCTGGTTGACAGTGCCTGGTGACGATGAAGGCGGCCCAT 3967
QY 130 CGGAGCGCGGACCTCGAGGCTATCGCGGAGCGTACCGCGAGATGGTGGCAGATCAC 189
DB 3966 GGTGAGCGCGCGCGCGCGCTTGGCGGACCGACCGCGCGCGCGCGCGCGCGAC 3907
QY 190 GAAATCGACGACGACACACGCGGAATCGCGCTGTTGTCGATGTCATCTCGACGATGAG 249
DB 3906 CGGAGGTGCGCGGGGTAGCCGCTGCTCGGTTGAGGCGGTTGGGAGGTTGATGAG 3847
QY 250 CGCGGCTTGAGGCGGGGATGAAGCTCGGCTGGCATCGGTATCACTTCCCCGACGAA 309

```

```

Db      3846  CGGAGGATGACGTGGATGGGGGTTGTGGTGTGGCGCCGAGTCAGCCTCGACAGC 3787
Qy      310  GACAGC 315
Db      3786  GATATC 3781

```

```

RESULT 15
US-09-252-991A-6325
; Sequence 6325, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6325
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6325

```

```

Query Match      12.3%; Score 39.4; DB 4; Length 1470;
Best Local Similarity 54.5%; Pred.No. 0.19;
Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy      1  TTGACCCACAAGCGCACTAAGCGGAGCCAGCCATCGCCGAGGGGTCAAGCCCGCGT 60
Db      979  TTTCACCGCGAGATCGCCGAGGGCCAGCGGCGATCGGCGCCGAGCCCTTGAGGGT 1038

Qy      61  CGGAATCGCGTTGGCGGCAACATGTTGGCGGCGGACGTTCCGTCGCGGAGCAGCGC 120
Db      1039  CGACAGCAGCGGGAGGCCAGCTTGGGTACTGGCGGTGCGCCAGTTCGGCTCCTTGCC 1098

Qy      121  CGGCGCCCAACGGCAGCGCGACCTCG 145
Db      1099  GGGCGCCGACCAGCAGCAGCGGCTCG 1123

```

Search completed: February 10, 2005, 06:44:45  
Job time : 96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:46:41 ; Search time 2098 Seconds  
(without alignments)  
5575.383 Million cell updates/sec

Title: US-10-647-089-93  
Perfect score: 321  
Sequence: 1 ttgacccacagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49.8	15.5	925	9	CNS0091P
2	46.8	14.6	627	7	CF481544
3	45.8	14.6	792	7	CF487459
4	45.2	14.1	626	6	CD927090
5	45	14.0	925	9	CNS0091P
6	44.8	14.0	1101	9	CNS017SY
7	43.6	13.6	532	9	P233L
8	43.6	13.6	623	9	P305L
9	43	13.4	478	6	CB029421
10	42.8	13.3	1075	8	BZ574121
11	42.8	13.3	1331	8	BZ576260
12	42.8	13.3	1348	8	BZ574088
13	42.6	13.3	932	9	CNS0072Q
14	42.4	13.2	514	6	CA248787
15	42.4	13.2	809	9	CC526073
16	42.4	13.2	906	9	CG178817
17	42.4	13.2	912	9	CG178819
18	42.4	13.2	958	9	CG084181
19	42.4	13.2	960	9	CC667226
20	42.2	13.1	734	7	CF482703
21	42	13.1	323	4	BI776880
22	42	13.1	429	5	BU983832
23	42	13.1	530	6	CD895551
24	42	13.1	638	6	CD902568

C 25	42	13.1	658	5	BQ805102
C 26	42	13.1	659	6	CD885865
C 27	42	13.1	747	9	CNS02PY0
28	41.8	13.0	1673	5	BU907501
29	41.6	13.0	611	6	CD938304
C 30	41.6	13.0	741	7	CF486940
31	41.6	13.0	846	9	CNS010RJ
C 32	41.2	12.8	780	8	BH370998
33	41.2	12.8	844	9	CNS0052P
C 34	41.2	12.8	1101	9	CNS01720
C 35	41.2	12.8	1237	7	CK211563
36	41	12.8	400	1	AA443087
37	40.8	12.7	673	6	CA221306
38	40.8	12.7	942	5	BU534531
39	40.6	12.6	1254	9	AG057481
40	40.6	12.6	1542	9	AG032943
41	40.4	12.6	704	9	AG060135
42	40.4	12.6	802	9	AG132220
C 43	40.2	12.5	537	4	BI804974
44	40.2	12.5	839	9	CNS004NB
45	40.2	12.5	951	6	CD492951

## ALIGNMENTS

RESULT 1  
CNS0091P  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
AL053013  
VERSION  
AL053013.1 GI:4934461  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 925)  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
JOURNAL  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
location/Qualifiers  
1. 925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

Query Match 15.5%; Score 49.8; DB 9; Length 925;  
Best Local Similarity 14.7%; Pred. No. 0.024;







Query Match	14.0%;	Score 45;	DB 9;	Length 925;	
Best Local Similarity	13.8%;	Pred. No. 0.4;			
Matches	40;	Conservative 132;	Mismatches 117;	Indels 0;	Gaps 0;
QY	22	CGCCAGCAGCATCGCGCAGGAGCTCAACGCCGCCGTCGGATCGGTTGGCGCGCAA	81		
Db	921	CGSCGCSBSCSSSMTSSSSBSCSCSSBSSSTSSMSSSSSSSSSSSSSGT	862		
QY	82	CATGGTTGGCGGCGGACGTTTCCTCCGCCGAGCAGCGCCGCCCAACGCGCAGCGCGAC	141		
Db	861	SSACVKCNASSCGCCGCMABCMCSSSSSSCGSASARGVKVRASGAGKRGSGSGAS	802		
QY	142	CTGAGGCTATCGCCGAGCTACGCCGAGTGATGGTGGCGCATCATCAGCAAAATCAGCAG	201		
Db	801	ASHSSSSACBSSSSSCBASCSWSSSSSSSSSSSSRSRGGGAGGSSSSSSSSSSASAG	742		
QY	202	GACACAGCCCAACTGGCGCTGTGTTCATGCATCTCAGCATGACGACGCGCGCTTGAG	261		
Db	741	SVSSASSSSSSSCSSVSCSVASSMSCSSBSSSSSASASSSSSSSSSASCSCCCTWS	682		
QY	262	CGCGGATGAAGCTCGCTGGCATCCGTATCACTTCCCGACAGAACCCG	310		
Db	681	CSCSTASMSAARSSSSSSSSSCSSSMSSASSSSSSSSSSSSSSSSSSSSSS	633		
RESULT 6					
CNS017SY		1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
Source					
ORIGIN					
Query Match	14.0%;	Score 44.8;	DB 9;	Length 1101;	
Best Local Similarity	18.2%;	Pred. No. 0.45;			
Matches	51;	Conservative 128;	Mismatches 99;	Indels 2;	Gaps 1;
QY	6	CCACAGCGCACTAAACGCCAGCCAGCCATCGCGCAGGCTCAAGCCCGCGTGGAA	65		
Db	699	CNAADCGCCAKWACSSSSSSSSSSSASTSSRSGRGMVSCACGSGSGASACGAS	758		

Db	482	ACGTGATCAAGCGCGCAAGGAGTGTCTTACGGCGCGCGATCCGAAC	531
RESULT 8			
LOCUS	P305L		
DEFINITION	Leishmania major Friedlin PAC P305 left end-sequence, similar to ECPHN P90227 Escherichia coli phn operon genes. . . . N=835, Prob=7.2e-59; SW:PHND_ECOLI P16682 PHOSPHONATES-BINDING. . . . N=480, Prob=1.4e-61, genomic_survey sequence.		
ACCESSION	AL160634		
VERSION	AL160634.1	GI:7259018	
KEYWORDS	GSS.		
SOURCE	Leishmania major		
ORGANISM	Leishmania major		
REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.		
AUTHORS	1 (bases 1 to 623)		
TITLE	Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F.		
JOURNAL	A physical map of the Leishmania major Friedlin genome		
MEDLINE	Genome Res. 8 (2), 135-145 (1998)		
PUBMED	98146435		
REFERENCE	9477341		
AUTHORS	2 (bases 1 to 173)		
TITLE	Barrell, B.G.		
JOURNAL	Submitted (14-MAR-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and alicat@sanger.ac.uk		
COMMENT	see http://www.ebi.ac.uk/parasites/leish.html Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/L_major/ The primer sequence can be obtained from alicat@sanger.ac.uk.		
FEATURES	Location/Qualifiers		
source	1. .623		
ORIGIN	/organism="Leishmania major"		
Query Match	13.6%; Score 43.6; DB 9; Length 623;		
Best Local Similarity	53.5%; Pred. No. 0.92;		
Matches	91; Conservative 0; Mismatches 79; Indels 0; Gaps 0;		
QY	143 TCGAGGCTATCCGCGGAGCGTACGCGGAGAGTGTGGCGACATCACAGAAATCGACGACG 202		
Db	347 TGGAGCGGTGATCGCGCAACGCGGAGGTGTGCGCTCGGTGATCGACAGACGGCA 406		
QY	203 ACACAGCGGAATGCGCGTGTGTTCGATGTCATCTCGACGATGACAGCGCGCGTTGAGG 262		
Db	407 ACCCGGGCTACTGTTGCTGCTGCTGATCGTCAGCAAGGACAGCGACCTCAAGTCGTCGACG 466		
QY	263 CGGGGATGAGCTCGGCTGGCATCGTATCACTTCCCGGACGACCCGAC 312		
Db	467 ACCTGATCAAGCGCGGCAAGGAGTGTCTTACGGCGCGCGATCCGAAC 516		
RESULT 9			
LOCUS	CB029421		
DEFINITION	TgESTzyd65f02.y2 TgrH Tachyzoite Norm 5 cDNA Library Toxoplasma gondii cDNA clone TgESTzyd65f02.y2 5', mRNA sequence.		
ACCESSION	CB029421		
VERSION	CB029421.1	GI:27725793	
KEYWORDS	EST.		
SOURCE	Toxoplasma gondii		
ORGANISM	Toxoplasma gondii		
REFERENCE	1 (bases 1 to 478)		
AUTHORS	Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajioaka, J.A., White, M., Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Teagareishvili, R., Ronko, I., Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.		
TITLE	Toxoplasma EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Clifton, S. Toxoplasma EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: toxo@watson.wustl.edu Contact David Sibley (toxoe@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability Seq primer: -40RP from Gibco High quality sequence stop: 439.		
FEATURES	Location/Qualifiers		
source	1. .478		
ORIGIN	/organism="Toxoplasma gondii"		
Query Match	13.4%; Score 43; DB 6; Length 478;		
Best Local Similarity	49.8%; Pred. No. 1.3;		
Matches	109; Conservative 0; Mismatches 110; Indels 0; Gaps 0;		
QY	63 GAATCGCGTTGGCGGCAACATGTTGGCGCGGCGAGTTCGTCGCCGAGCAGCGCG 122		
Db	63 GAAACCGGCTCGAGTCGGGAGCAGTCTGTACACTCGAGCGGAGGTCGCGAGCGGATCCG 122		
QY	123 CGCCCAAGCGGCGGCGACCTCGAGGCTATCCCGGAGCGGTACCGGAGATGTGGCGAC 182		
Db	123 AGAAGAGCGACAGCTTCAGTTCAGGAGGTCGCGGAGAGCAGCGGAGGAGTAGCGGAG 182		
QY	183 ATCACAGAAATCGACGACACAGCGGAACTGGGCGTCTTTCGATGTCATCTCGACGA 242		
Db	183 GCTCCAGGCTGACAGCCGAGATCCGGGCGAGACAGAGAAGTTCGACGCGCTCTGTGA 242		
QY	243 TGAGCAGCGCGCGTTGAGCGGGGATGAAGTTCGGCTG 281		
Db	243 CGAGCAACTCGAGTTCAGGAAGCGGAAGAAGCTCGCGG 281		
RESULT 10			
LOCUS	BZ574121/c		
DEFINITION	1075 bp DNA linear		
ACCESSION	BZ574121		
VERSION	GSS 17-DEC-2000		

```

DEFINITION msh2_3520.x1 msh Pseudomonas aeruginosa genomic clone msh2_3520,
             genomic survey sequence.
ACCESSION  BZ574121
VERSION    BZ574121.1 GI:27209182
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1075)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol. (2002) In press
COMMENT    Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES   Location/Qualifiers
            source          1..1075
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="MSH"
                        /db_xref="taxon:287"
                        /clone_lib="msh2_3520"
                        /notes="Environmental isolate. Whole genomic shotgun
                        library."
ORIGIN
Query Match      13.3%; Score 42.8; DB 8; Length 1075;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY  48 CAACGCCCGCTCGGAATCGCTTGGGGGCAACATGTTGGCGCGCGACGTTCCGTC 107
Db  392 CCAGTCTCGAGGACGATATCGGCTCGCGGCGGACAGATGTTGTTGG 333
QY  108 CGCGAGCAGCGCGCGCCCAACGGCAGCGCGACCTCGAGGCTATCCGCGAGCGTACGC 167
Db  332 TGGACATCATCAGCGCGCGCTTCAGCTGGGCTTCGAGGGTCAGCGGAACGTGACGG 273
QY  168 CGAGATGGTGGGACATCACAGAAATCGACGACGACACAGCGCACTGGCGCTGTTGTC 227
Db  272 CCATCTGGTCACCGTCAAGTCGGCGTTGTACGCGGCGCAGACCGCGGTGCAGCTGA 213
QY  228 GATGCATCTCGACGATGAGCAGCGCC 253
Db  212 TCGCTTGCCTTCGATGAGACCGGC 187

RESULT 11
BZ576260/c
LOCUS      1331 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_4849.y2 msh Pseudomonas aeruginosa genomic clone msh2_4849,
             genomic survey sequence.
ACCESSION  BZ576260
VERSION    BZ576260.1 GI:27211321
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1331)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library

```

```

JOURNAL      J. Bacteriol. (2002) In press
COMMENT      Contact: Chris K. Raymond
             Genome Center
             University of Washington
             Box 352145, Seattle, WA 98105-2145, USA
             Tel: 2062216954
             Fax: 2066857244
             Email: craymond@u.washington.edu
             Class: shotgun.
FEATURES     Location/Qualifiers
            source          1..1331
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="MSH"
                        /db_xref="taxon:287"
                        /clone_lib="msh2_4849"
                        /notes="Environmental isolate. Whole genomic shotgun
                        library."
ORIGIN
Query Match      13.3%; Score 42.8; DB 8; Length 1331;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY  48 CAACGCCCGCTCGGAATCGCTTGGGGGCAACATGTTGGCGCGCGACGTTCCGTC 107
Db  348 CCAGTCTCGAGGACGATATCGGCTCGCGGCGGACAGATGTTGTTGG 289
QY  108 CGCGAGCAGCGCGCGCCCAACGGCAGCGCGACCTCGAGGCTATCCGCGAGCGTACGC 167
Db  288 TGGACATCATCAGCGCGCGCTTCAGCTGGGCTTCGAGGGTCAGCGGAACGTGACGG 229
QY  168 CGAGATGGTGGGACATCACAGAAATCGACGACGACACAGCGCACTGGCGCTGTTGTC 227
Db  228 CCATCTGGTCACCGTCAAGTCGGCGTTGTACGCGGCGCAGACCGCGGTGCAGCTGA 169
QY  228 GATGCATCTCGACGATGAGCAGCGCC 253
Db  168 TCGCTTGCCTTCGATGAGACCGGC 143

RESULT 12
BZ574088/c
LOCUS      1348 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_3505.y3 msh Pseudomonas aeruginosa genomic clone msh2_3505,
             genomic survey sequence.
ACCESSION  BZ574088
VERSION    BZ574088.1 GI:27209149
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1348)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol. (2002) In press
COMMENT    Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES     Location/Qualifiers
            source          1..1348
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="MSH"

```

```

/db_xref="taxon:287"
/clone="msh2.3505"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN
Query Match      13.3%; Score 42.8; DB 8; Length 1348;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 48 CAACGCCCGGTCGGATCGGTTGGCGGCAACATGTTGGCGCGACGTTCCGTC 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 CCACGTCTCGGACGACGATGATCGGCTCGCCGTTGGCGGCGACAGGATGTTGG 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 CGCGAGCAGCGCGCGCCCAACGCGCAGCGGACCTCGAGGCTATCCCGGCGTACGC 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 TGGACATCATCAGCGCGCGCGTTCAGCTGGGCTCGAGGTCAGCGAAGTGGACGG 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 CGAGATGTGGCGCATCACACGAAATCGACGACACACGCGCAACTGGGCTGTTGTC 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 CCATCTGTACCGTCGAGTCGGGTTGTACGCGCGCGACGACGCGGTGCAGCTGA 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 228 GATGCATCTCGACGATGAGCGGCC 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 TCGCCTTGCTTCGATGAGGACCGGC 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
CNS0072Q      932 bp DNA linear GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION      fly), genomic survey sequence.
ACCESSION      AL066742
VERSION      AL066742.1 GI:4945205
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 932)
AUTHORS      Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT      - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mamoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              P1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
              Location/Qualifiers
              1..932
              /organism="Drosophila melanogaster"
              /mol_type="genomic DNA"
              /db_xref="taxon:7227"
              /clone="BAC14B09"
              /clone_lib="RPCI-98"
              /note="end : T7"

ORIGIN
Query Match      13.2%; Score 42.4; DB 6; Length 514;
Best Local Similarity 49.3%; Pred. No. 1.9;
Matches 109; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 92 CGGCGACGTTTCGTCGCCGAGCAGCGCGGCCCAACGCGCGACCTCGAGCTA 151

```

```

Query Match      13.3%; Score 42.6; DB 9; Length 932;
Best Local Similarity 37.2%; Pred. No. 1.6;
Matches 51; Conservative 36; Mismatches 50; Indels 0; Gaps 0;

QY 22 GCCACGCCAGCATCGCGCAGGGCTCAACGCCCGCGTCGGAATCGGTCGTTGGGGCAA 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 CSCCCSCCCSCGSSSGCGSGGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGSG 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 CATGTTGGCGCGGACGTTCCGTCGCCGAGCAGCGCGCCCAACGCGCAGCGCGAC 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 CGSGSGCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 CTCGAGGCTATCCGCG 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 SSCGCGSGCGSGSCG 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
CA248787      514 bp mRNA linear EST 25-SEP-2003
LOCUS      SCSBFL1104G10.9 FL1 Saccharum officinarum cDNA clone SCSBFL1104G10
DEFINITION      5' mRNA sequence.
ACCESSION      CA248787
VERSION      CA248787.1 GI:35330788
KEYWORDS      EST.
SOURCE      Saccharum officinarum
ORGANISM      Saccharum officinarum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
              complex.
REFERENCE      1 (bases 1 to 514)
AUTHORS      Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
TITLE      The libraries that made SUCEST
JOURNAL      Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT      Contact: Arruda P
              Centro de Biologia Molecular e Engenharia Genetica
              Universidade Estadual de Campinas
              Caixa Postal 6010, 13083-970, Campinas SP, Brazil
              Tel: 55 19 3788 1137
              Fax: 55 19 3788 1089
              Email: parruda@unicamp.br
              Clone distribution: clone distribution information can be found
              through the Brazilian Clone Collection Center (BCCC) at
              http://www.bcccenter.fcav.unesp.br
              Plate: 104 row: G column: 10
              Seq primer: T7 Promoter Primer.
              Location/Qualifiers
              1..514
              /organism="Saccharum officinarum"
              /mol_type="mRNA"
              /db_xref="taxon:4547"
              /clone="SCSBFL1104G10"
              /lab_host="DH10B"
              /clone_lib="FL1"
              /note="Organ: Inflorescence at beginning of development
              (1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
              An unidirectional cDNA library generated from
              (inflorescence at beginning of development (1cm-long)).
              cDNA was prepared from polyA+ mRNA using SuperScript
              Plasmid System Kit (Invitrogen). The double-strand cDNAs
              were fractionated in a sepharose CL-2B 40cm-columns and
              fragments sizing between 0.8 and 1.5 Kb were
              directionally cloned into the vector. Details
              of each
              source of RNA and library construction can be obtained at
              http://sucest.lad.ic.unicamp.br/public"

ORIGIN
Query Match      13.2%; Score 42.4; DB 6; Length 514;
Best Local Similarity 49.3%; Pred. No. 1.9;
Matches 109; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 92 CGGCGACGTTTCGTCGCCGAGCAGCGCGGCCCAACGCGCGACCTCGAGCTA 151

```

Search completed: February 10, 2005, 06:43:13  
Job time : 2108 secs

```
Db      293  CGCCCGGAGCGCGCGGAGATCAACCTCTCCCTGCCCGGAGCGCTTCGAGCGCG 352
QY      152  TCCGCGGAGCGTACCGGAGATGGTGGGACATCATCAGCAAAATCGACGACGACGCG 211
Db      353  TGGGCGCGCCCAAGCGGCGCTGGAGGCGCGCTGCCCGCGCGTGTGTCTCGCGCGAGC 412
QY      212  AACTGGCGCTGTGTTCGATGCATCTCGACGATGAGCAGCGCGCGTTCGAGGCGGGGATGA 271
Db      413  TCCTGGCCCTCGCCGCGGCGGACCTCTGTGCCATCTCTGGCGGCGCACGGGTCCCGGTG 472
QY      272  AGCTCGGTGGCATCGTATCATCTTCCCGCGAAGCCCGAC 312
Db      473  CGCTGGCGCGCGNGACTCCGCCGCTCCGACGCGCGCGAC 513
```

## RESULT 15

CC626073/c

LOCUS

DEFINITION OGUHF43TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma0439H14,  
genomic survey sequence.

ACCESSION CC626073

VERSION CC626073.1 GI:31995181

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics

JOURNAL Unpublished (2002)

COMMENT

Other\_GSSs: OGUHF43TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

## FEATURES

source

1..809

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBma0439H14"

/clone\_lib="ZM 0.7 1.5 KB"

/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

## ORIGIN

Query Match 13.2%; Score 42.4; DB 9; Length 809;

Best Local Similarity 53.7%; Pred. No. 1.8;

Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 96 CGACGTTCCGTCGCCGAGCAGCGCGCCCAACGCGACGCGGACCTCGAGGCTATCCG 155

Db 309 CGCGGATTCGGCGGGGCGCTGCGGTTCGCGACTGCTCGCGACGCGGAGCCCGACGA 250

QY 156 CCGAGCGTACCGCGAGATGGTGGGACATCACACGAATTCGACGACGACGCGGAACT 215

Db 249 GTTGGAGTGGCGCTGATGGTGGCGAGCTGCACAGAGCTCAAGGGGAGCCTCCCGAGGC 190

QY 216 GCGCTGTGTTTCGATGCATCTCGACGATGACGAGCGCGCGCTTG 259

Db 189 TGAGCGTCACTTCAGGAGCTCTCGGACGAGGAGCCCTCTCTCG 146

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:44:31 ; Search time 335 Seconds  
(without alignments)  
5030.042 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321

Sequence: 1 ttgaccacacagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_23Sep04.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	321	100.0	324	4	Aaf24416	Aaf24416 M tubercu
2	321	100.0	110000	4	AAI199882_29	Continuation (30 o
3	321	100.0	110000	4	AAI199883_29	Continuation (30 o
4	42.8	13.3	4200	4	AAI199883_29	Continuation (30 o
5	42.8	13.3	4200	8	ACA19522	ACA19522 Pseudomon
6	42.8	13.3	4212	11	ABD06144	Abd06144 Pseudomon
7	42.8	13.3	4242	11	ABD08093	Abd08093 Pseudomon
8	42.8	13.3	10232	8	ACA37617	ACA37617 Pseudomon
9	41.6	13.0	996	8	ACA23398	ACA23398 Pseudomon
10	41	12.8	400	6	ABL1990	ABL1990 Human ova
11	40.4	12.6	951	8	ACA38286	ACA38286 Pseudomon
12	40.4	12.6	954	4	AH51985	AH51985 Mycobacte
13	40.4	12.6	954	8	ACA40430	ACA40430 Pseudomon
14	40.4	12.6	45190	6	ABX09144	ABX09144 Mycobacte
15	40.4	12.6	110000	4	AAI199882_11	Continuation (12 o
16	40.4	12.6	110000	4	AAI199883_11	Continuation (12 o
17	40.2	12.5	732	11	ABD07694	Abd07694 Pseudomon
18	40.2	12.5	945	11	ABD07563	Abd07563 Pseudomon
19	40.2	12.5	1602	11	ABD07778	Abd07778 Pseudomon
20	39.8	12.4	594	11	ABD13439	ABD13439 Pseudomon
21	39.8	12.4	966	11	ABD13292	ABD13292 Pseudomon

22	39.8	12.4	1179	11	ABD13413	Abd13413 Pseudomon
23	39.6	12.3	1512	8	ABX71825	Abx71825 DNA encod
24	39.6	12.3	4509	8	ABX71813	Abx71813 Schizochy
25	39.6	12.3	4512	8	ABX71823	Abx71823 DNA encod
26	39.6	12.3	4767	3	AAA71571	Aaa71571 S. aggreg
27	39.4	12.3	1470	11	ABD07721	Abd07721 Pseudomon
28	39.4	12.3	1473	11	ABD07540	Abd07540 Pseudomon
29	39.4	12.3	1488	4	AA54225	Aa54225 Pseudomon
30	39.4	12.3	1488	8	ACA42505	Aca42505 Prokaryot
31	39.4	12.3	1512	11	ABD07615	Abd07615 Pseudomon
32	39.2	12.2	1320	8	ABX56068	Abx56068 M. echino
33	38.8	12.1	408	11	ABD02051	Abd02051 Pseudomon
34	38.8	12.1	603	8	ACA26989	ACA26989 Prokaryot
35	38.8	12.1	801	11	ABD02094	Abd02094 Pseudomon
36	38.8	12.1	1356	11	ABD02026	Abd02026 Pseudomon
37	38.8	12.1	1434	8	ACA35463	Aca35463 Prokaryot
38	38.8	12.1	2955	11	ABD02080	Abd02080 Pseudomon
39	38.8	12.1	3150	11	ABD02123	Abd02123 Pseudomon
40	38.8	12.1	109519	5	AA508693	Aa508693 Micromono
41	38.6	12.0	2379	4	AA54231	Aa54231 Pseudomon
42	38.6	12.0	2379	8	ACA42520	Aca42520 Prokaryot
43	38.6	12.0	2379	8	ABT14638	Abt14638 Pseudomon
44	38.4	12.0	1371	11	ABD15350	Abd15350 Pseudomon
45	38.4	12.0	2181	11	ABD15138	Abd15138 Pseudomon

ALIGNMENTS

RESULT 1  
AAF24416  
ID AAF24416 standard; DNA; 324 BP.

AC AAF24416;

DT 24-APR-2001 (first entry)

DE M tuberculosis Rv2653c gene.

XX Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c; Rv1037c;  
KW Rv2346c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c; Rv3445c; Rv3890c;  
KW Rv3891c; Rv3904c; Rv3905c; ds.

OS Mycobacterium tuberculosis.

PN WO200104151-A2.

XX 18-JAN-2001.

PF 13-JUL-2000; 2000WO-DK000398.

PR 13-JUL-1999; 99DK-00001020.

PA 15-JUL-1999; 99US-0144011P.

XX (STAT-) STATENS SERUM INST.

PI Andersen P, Skjot R;

DR WPI; 2001-091923/10.

XX P-FSDB; AAB35226.

XX Immunizing against and diagnosis of tuberculosis.

PS Claim 28; Page 71; 80pp; English.

CC The present invention provides the protein and coding sequences for  
members of the esat-6 gene family from Mycobacterium tuberculosis. These  
proteins include Rv0287, Rv1037c, Rv2346c, Rv2348c, Rv2653c,  
Rv2654c, Rv3020c, Rv3444c, Rv3890c, Rv3891c, Rv3904c and  
Rv3905c. These can be used to produce vaccines against, and in the  
diagnosis of, tuberculosis (TB) infection. The present sequence is one of  
the coding sequences of the invention

XX SQ Sequence 324 BP; 69 A; 112 C; 100 G; 43 T; 0 U; 0 Other;

Query Match 100.0%; Score 321; DB 4; Length 324;  
Best Local Similarity 100.0%; Pred. No. 5.9e-70;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGACCCACAGCGCACTAAACGCCAGCCAGCCATCGCCGAGGCTCAACGCCCGCGT 60  
DB 1 TTGACCCACAGCGCACTAAACGCCAGCCAGCCATCGCCGAGGCTCAACGCCCGCGT 60

QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGACGTTCCGTCGCGAGCAGCGC 120  
DB 61 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGACGTTCCGTCGCGAGCAGCGC 120

QY 121 CGCGCCCAACGGCAGCCGACCTCGAGGCTATCGCCGAGCGTACGCCGAGATGGTGGCG 180  
DB 121 CGCGCCCAACGGCAGCCGACCTCGAGGCTATCGCCGAGCGTACGCCGAGATGGTGGCG 180

QY 181 ACATCACAGAAATCGACGACACAGCCGAACTGGCGCTGTTGCGATGCTCTCGAC 240  
DB 181 ACATCACAGAAATCGACGACACAGCCGAACTGGCGCTGTTGCGATGCTCTCGAC 240

QY 241 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTCCCC 300  
DB 241 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTCCCC 300

QY 301 GACGAACCCGACAGCAACAG 321  
DB 301 GACGAACCCGACAGCAACAG 321

RESULT 2  
AAI99682\_29/c  
Continuation (30 of 45) of AAI99682 from base 2900001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

Fragment Name	Begin	End
AAI99682_00	1	110000
AAI99682_01	100001	210000
AAI99682_02	200001	310000
AAI99682_03	300001	410000
AAI99682_04	400001	510000
AAI99682_05	500001	610000
AAI99682_06	600001	710000
AAI99682_07	700001	810000
AAI99682_08	800001	910000
AAI99682_09	900001	1010000
AAI99682_10	1000001	1110000
AAI99682_11	1100001	1210000
AAI99682_12	1200001	1310000
AAI99682_13	1300001	1410000
AAI99682_14	1400001	1510000
AAI99682_15	1500001	1610000
AAI99682_16	1600001	1710000
AAI99682_17	1700001	1810000
AAI99682_18	1800001	1910000
AAI99682_19	1900001	2010000
AAI99682_20	2000001	2110000
AAI99682_21	2100001	2210000
AAI99682_22	2200001	2310000
AAI99682_23	2300001	2410000
AAI99682_24	2400001	2510000
AAI99682_25	2500001	2610000
AAI99682_26	2600001	2710000
AAI99682_27	2700001	2810000
AAI99682_28	2800001	2910000
AAI99682_29	2900001	3010000
AAI99682_30	3000001	3110000
AAI99682_31	3100001	3210000
AAI99682_32	3200001	3310000
AAI99682_33	3300001	3410000
AAI99682_34	3400001	3510000
AAI99682_35	3500001	3610000

WP AAI99682\_36 3600001 3710000  
WP AAI99682\_37 3700001 3810000  
WP AAI99682\_38 3800001 3910000  
WP AAI99682\_39 3900001 4010000  
WP AAI99682\_40 4000001 4110000  
WP AAI99682\_41 4100001 4210000  
WP AAI99682\_42 4200001 4310000  
WP AAI99682\_43 4300001 4410000  
WP AAI99682\_44 4400001 4411529

Query Match 100.0%; Score 321; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 9.9e-70;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGACCCACAGCGCACTAAACGCCAGCCAGCCATCGCCGAGGCTCAACGCCCGCGT 60  
DB 76906 TTGACCCACAGCGCACTAAACGCCAGCCAGCCATCGCCGAGGCTCAACGCCCGCGT 76847

QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGACGTTCCGTCGCGAGCAGCGC 120  
DB 76846 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGACGTTCCGTCGCGAGCAGCGC 76787

QY 121 CGCGCCCAACGGCAGCCGACCTCGAGGCTATCGCCGAGCGTACGCCGAGATGGTGGCG 180  
DB 76786 CGCGCCCAACGGCAGCCGACCTCGAGGCTATCGCCGAGCGTACGCCGAGATGGTGGCG 76727

QY 181 ACATCACAGAAATCGACGACACAGCCGAACTGGCGCTGTTGCGATGCTCTCGAC 240  
DB 76726 ACATCACAGAAATCGACGACACAGCCGAACTGGCGCTGTTGCGATGCTCTCGAC 76667

QY 241 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTCCCC 300  
DB 76666 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTCCCC 76607

QY 301 GACGAACCCGACAGCAACAG 321  
DB 76606 GACGAACCCGACAGCAACAG 76586

RESULT 3  
AAI99683\_29/c  
Continuation (30 of 44) of AAI99683 from base 2900001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683

Fragment Name	Begin	End
AAI99683_00	1	110000
AAI99683_01	100001	210000
AAI99683_02	200001	310000
AAI99683_03	300001	410000
AAI99683_04	400001	510000
AAI99683_05	500001	610000
AAI99683_06	600001	710000
AAI99683_07	700001	810000
AAI99683_08	800001	910000
AAI99683_09	900001	1010000
AAI99683_10	1000001	1110000
AAI99683_11	1100001	1210000
AAI99683_12	1200001	1310000
AAI99683_13	1300001	1410000
AAI99683_14	1400001	1510000
AAI99683_15	1500001	1610000
AAI99683_16	1600001	1710000
AAI99683_17	1700001	1810000
AAI99683_18	1800001	1910000
AAI99683_19	1900001	2010000
AAI99683_20	2000001	2110000
AAI99683_21	2100001	2210000
AAI99683_22	2200001	2310000
AAI99683_23	2300001	2410000
AAI99683_24	2400001	2510000
AAI99683_25	2500001	2610000
AAI99683_26	2600001	2710000
AAI99683_27	2700001	2810000
AAI99683_28	2800001	2910000



WP AAI99683\_29 2900001 3010000  
 WP AAI99683\_30 3000001 3110000  
 WP AAI99683\_31 3100001 3210000  
 WP AAI99683\_32 3200001 3310000  
 WP AAI99683\_33 3300001 3410000  
 WP AAI99683\_34 3400001 3510000  
 WP AAI99683\_35 3500001 3610000  
 WP AAI99683\_36 3600001 3710000  
 WP AAI99683\_37 3700001 3810000  
 WP AAI99683\_38 3800001 3910000  
 WP AAI99683\_39 3900001 4010000  
 WP AAI99683\_40 4000001 4110000  
 WP AAI99683\_41 4100001 4210000  
 WP AAI99683\_42 4200001 4310000  
 WP AAI99683\_43 4300001 4403765

Query Match 100.0%; Score 321; DB 4; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-70;  
 Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGACCCACAGCGCACTAAAGCCAGCCAGCCATCGCGGAGGCTCAACGCCCGCGGT 60  
 DB 71654 TTGACCCACAGCGCACTAAAGCCAGCCAGCCATCGCGGAGGCTCAACGCCCGCGGT 71595

QY 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGGACGTTCCGTCGCGGAGCAGCGC 120  
 DB 71594 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGGACGTTCCGTCGCGGAGCAGCGC 71535

QY 121 CGCGCCCAAGCGCAGCGCACTCGAGGCTATCCGCCGAGCGTACGCCGAGATGTCGCG 180  
 DB 71534 CGCGCCCAAGCGCAGCGCACTCGAGGCTATCCGCCGAGCGTACGCCGAGATGTCGCG 71475

QY 181 ACATCACAGCAATCGACGACGACACGACGCACTGCGGCTGTTGTCGATGATCTCGAC 240  
 DB 71474 ACATCACAGCAATCGACGACGACACGACGCACTGCGGCTGTTGTCGATGATCTCGAC 71415

QY 241 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGTCGATCCGTCATCCTCC 300  
 DB 71414 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGTCGATCCGTCATCCTCC 71355

QY 301 GACGAACCGCAGCAAAACAG 321  
 DB 71354 GACGAACCGCAGCAAAACAG 71334

RESULT 4  
 AAS51542/c  
 ID AAS51542 standard; DNA; 4200 BP.  
 AC AAS51542;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Pseudomonas aeruginosa DNA for cellular proliferation protein #127.  
 KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
 KW antibacterial; drug design.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US009180.  
 XX  
 PR 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207272P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.  
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 XX Yamamoto RT, Xu HH;  
 PI WPI; 2001-611495/70.  
 XX P-PSDB; AAU33683.  
 DR  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 PS Claim 27; SEQ ID NO 4124; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence encodes an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 4200 BP; 846 A; 1329 C; 1304 G; 721 T; 0 U; 0 Other;  
 Query Match 13.3%; Score 42.8; DB 4; Length 4200;  
 Best Local Similarity 50.5%; Pred. No. 0.42;  
 Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 48 CAACGCCCGCGTGGGATCGGTTGGCGCGCAACATGTTGGCGCGGAGCTTCCGTC 107  
 DB 1519 CCAGTCTCGACGGAACGATGTCGGCTCGCCGTTGGCGCGCAGAGATGTTGTTGG 1460

QY 108 CGCGAGCAGCGCGCGCCCAACGCGCAGCGGACCTCGAGGCTATCCGCCGAGCGTACGC 167  
 DB 1459 TGGACATCATCAGCGCGCGCTTCCAGCTGGGCTCGAGGTCAGCGGACGTGGACGG 1400

QY 168 CGAGATGTCGCGACATCACACGAAATCGACGACGACACAGCGGAACTGGCGCTGTTGC 227  
 DB 1399 CCATCTGTCACCGTCGAAGTCGGCGTTGTACGCGCGCAGACCGCGGTGTCAGCTGGA 1340

QY 228 GATGCATCTCGACGATGAGCGGCC 253  
 DB 1339 TGGCCTTCCTTCGATGAGGACCGGC 1314

RESULT 5  
 ACA19522/c  
 ID ACA19522 standard; DNA; 4200 BP.  
 AC ACA19522;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #1179.  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN WO200277183-A2.  
 XX

PD 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362659P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wali D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 PI WPI; 2003-029926/02.  
 DR P-PSDB; ABU15652.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 14; SEQ ID NO 7392; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 4200 BP; 846 A; 1329 C; 1304 G; 721 T; 0 U; 0 Other;  
 Query Match 13.3%; Score 42.8; DB 8; Length 4200;  
 Best Local Similarity 50.5%; Pred. No. 0.42;  
 Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
 QY 48 CAACGCCCGCGTCCGGAATCGGTTGGCGGCAACATGTTGGCGCGCGAGTTCGTC 107  
 DB 1519 CCAGTCTCTGCGAGCGAAGATGATCGGTTGCCGTTGGCGGCGACAGATGTTGTTGG 1460  
 QY 108 CGCCGAGCAGCGCGCGCCCAACCGGACGCGCGCACTTCGAGGCTATCCGCCGAGCGTACGC 167  
 DB 1459 TGGACATCATCAGCGCGCGCGCTTCAGCTGGGCTCGAGGGTCAAGCGAAGTGGACGG 1400  
 QY 168 CGAGATGTTGGGACATCATCAGCAATTCAGACGACACAGCGCACTTCGCGCTGTTGTC 227  
 DB 1399 CCATCTGGTACCGTCAAGTTCGGCGTTTATCGCGCGCAGACCAAGCGGTTGACGTGGA 1340  
 QY 228 GATGCATCTCGACGATGACGAGCGCC 253  
 DB 1339 TCGCCTTGCTTCGATGAGGACCGGC 1314  
 RESULT 6  
 ABD06144/C  
 ID ABD06144 standard; DNA; 4212 BP.  
 XX AC ABD06144;  
 XX DT 29-JUL-2004 (first entry)  
 XX DE Pseudomonas aeruginosa polynucleotide #4748.  
 XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 KW antibacterial.  
 XX Pseudomonas aeruginosa.  
 OS US6551795-B1.  
 XX PN 22-APR-2003.  
 XX PD 18-FEB-1999; 99US-00252991.  
 XX PF 18-FEB-1998; 98US-0074788P.  
 PR 27-JUL-1998; 98US-0094190P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
 XX WPI; 2003-615309/58.  
 DR P-PSDB; ABO72573.  
 XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX Disclosure; SEQ ID NO 4748; 455pp; English.  
 XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
 CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
 CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
 CC of *Pseudomonas* species using biotech technology. Sequences ABD01397-  
 CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 XX  
 SQ Sequence 4212 BP; 850 A; 1338 C; 1311 G; 713 T; 0 U; 0 Other;  
 Query Match 13.3%; Score 42.8; DB 11; Length 4212;  
 Best Local Similarity 50.5%; Pred. No. 0.42;  
 Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
 QY 48 CAACGCCCGCGTCCGGAATCGGTTGGCGGCAACATGTTGGCGCGCGAGTTCGTC 107  
 DB 1531 CCAGTCTCTGCGAGCGAAGATGATCGGTTGCCGTTGGCGGCGACAGATGTTGTTGG 1472  
 QY 108 CGCCGAGCAGCGCGCGCCCAACCGGACGCGCGCACTTCGAGGCTATCCGCCGAGCGTACGC 167  
 DB 1471 TGGACATCATCAGCGCGCGCGCTTCAGTGGGCTCGAGGGTCAAGCGAAGTGGACGG 1412

## RESULT 7

Db 2865 TCGCCTTGCCCTTCGATGAGGACCGC 2890

## RESULT 8

ACA37617  
ID ACA37617 standard: DNA: 10232 BP.

AC ACA37617: XX

XX  
DT 19-JUN-2003 (first entry)  
XX

DE Prokaryotic essential gene #19274.  
XX

Antisense; ds; prokaryotic essential drug design: gene.

XX *Mycobacterium avium*.

XX PN WO200277183-A2

03-OCT-2002

21-MAR-2002: 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.  
XX

PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L. Zamudio C. Malone C. Ha

PI Wall D, Trawick JD, Carr GJ, Ya  
XX

DR WPI: 2003-029926/02.  
DR P-PSDE; ABU33747.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 25487; 17666bp; English.  
PS

CC the polypeptide; (5) producing the  
CC proliferation or the activity of a

CC proliferation; (7) identifying a c

CC required for proliferation, or that  
CC identifying a gene required for ce

CC pathway in which a proliferation-r  
CC or a gene on which the test compou

CC organism acts; (9) manufacturing a compound's activity; (11) a culture

product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://ftp.wipo.int/pub/published/pct_sequences)

SQ Sequence 10232 BP; 1443 A; 3537 C; 3647 G; 1605 T; 0 U; 0 Other;

Query Match	13.3%;	Score 42.8;	DB 8;	Length 10232;
Best Local Similarity	47.7%;	Pred. No. 0.45;		
Matches 125;	Conservative 0;	Mismatches 137;	Indels 0;	Gaps 0;
QY	20	AAGCGCAGCAGGCATCCGCGCAGGGCTCAACGCCGCCGCTCGGAATCGGTTGGGCGGC	79	
Db	1271	AACTGCTGACCGGGACGAGCCCGCGTGTATCGACCGTGGAGCAACCGTCCGCGCTGA	1330	
QY	80	AACATGTTGGCGCGGCCGAGCTTCCGTCGCGGAGCAGCGCCGCCGCCAACGGCAGGCGC	139	
Db	1331	CCGAGCCCCGACCCGCCGCCGGTTCGATCCCCCAGCGCTTTCGCCGAACACGTGCAGCGCA	1390	
QY	140	ACCTCGAGGCTATCCGCGCGAGCGGTACGCCGAGATGTTGGCGACATCACAGAAATCCAGC	199	
Db	1391	CCCCGACGCGTGGCGGTGACGTTCCGGGCGACCTCGTGACCTACGCCCCAGCTCGACG	1450	
QY	200	ACGACACAGCCGAACCTGGCGCTGTTTCGATGATCTTCGAGCATGACGAGCGCCGGTTG	259	
Db	1451	AGCGGTCAACCGGCTGGGCGCATCTGTCCGCGACCAACGCGTGGGCCCGGGCGCACTGCG	1510	
QY	260	AGCGGGGATGAAGCTCGGCTG	281	
Db	1511	TGGCGGTGATGTTCCCGCGCTG	1532	

RESULT 9  
ACA23398  
ID ACA23398 standard; DNA: 996 BP.

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #5055.

Antisense; ds; prokaryotic essential gene; cell proliferation;  
drug design; gene.  
KW

Borrelia cepacia.

AA  
PN  
WO200277183-A2.

03-OCT-2002.

21-MAR-2002: 2002WO-US009107.

21-MAR-2001: 2001US-00815242-XX  
PR

PK 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001: 2001US-0342923P

08-FEB-2002; 2002US-00072851.  
06-MAR-2002; 2002US-0362689P

XX  
PA (ELIT-) ELITPA PHARM INC

XX  
PT  
Wang  
I.  
Zhang  
C  
Ma

PI Wall D, Trawick JD, Car  
vv

WPI; 2003-029926/02.  
P-PSDB; ABU19528.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 14: SEQ ID NO 11268; 1766pp: English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://ftp.wipo.int/pub/published/pct_sequences)

Sequence 996 BP; 138 A; 320 C; 387 G; 151 T; 0 U; 0 Other;

every Match 13.0%: Score 41.6: DB 8: Length 996:

Best Local Similarity 50.8%; Pred. No. 0.73;

Mismatches 0; Conservative 127; Mismatches 119; Indels 4; Gaps 1;

29 CAGCCATCGCCGCAGGGCTCAACGCCCGCGTTCGGAATCGCGTTGGCGGCAACATGGTT 88

145 CTGCAACGGGCCGGCGTCCGCTCGATCGATGCGGGCGTCCCGCTGGCCGAGCGTTGGGAT 204

89 GGGCGGCGGACGTTCCGGTCCGGCCGAGGCGCGCGGCCCA-----CGGCAGCGGGGACCTC 144

205 CGCAGGGCCAGTTCGTACGGCGGTGGATGCCGGCA CGAGGCGGTGGTGGCGACGGC 264

145 GAGGCTATTCCGCCCGGGGTACGCCCGGCATGCTGGCGCATCACCACATCGACCGACCGAC 204

25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

— — — — —

## RESULT 10

ABL81990

ID ABL81990 standard; cDNA; 400 BP.

AC ABL81990;

```

XX DT 17-MAY-2002 (first entry)
XX AC Human ovarian cancer related cDNA clone SEQ ID NO:4968.
XX DE Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX KW Homo sapiens.
XX OS WO200192581-A2.
XX PN 06-DEC-2001.
XX PD 29-MAY-2001; 2001WO-US017756.
XX PF 26-MAY-2000; 2000US-0207484P.
XX PR (CORI-) CORIXA CORP.
XX PA Algate PA, Harlocker SL, Jones R;
XX PI WPI; 2002-122075/16.
XX DR Composition for therapy and diagnosis of ovarian cancer comprising
XX PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX PT polypeptide, antibody specific to polypeptide or T cell expressing
XX PT polypeptide.
XX PS Claim 1; SEQ ID NO 4968; 489pp; English.
XX PS The present invention describes a composition (I) comprising: carriers
XX CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
XX CC from the 10912 nucleotide sequences as given in ASL77023 to ABL87934.
XX CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
XX CC or antigen presenting cells that express (II). (I) has cytostatic
XX CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
XX CC detecting ovarian cancer in a patient's biological sample preferably
XX CC serum or ovarian tissue. The method comprises contacting a biological
XX CC sample from a patient with (IV), detecting the amount of polynucleotide
XX CC hybridising to (IV) and comparing the amount to a predetermined cutoff
XX CC value and thereby detecting ovarian cancer in the patient, where the
XX CC amount of polynucleotide hybridising to (IV) is detected preferably by
XX CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
XX CC useful for stimulating and/or expanding T cells specific for an ovarian
XX CC tumour protein comprising contacting T cells with (III) or (II). (III) is
XX CC useful in design and preparation of ribozyme molecules for inhibiting
XX CC expression of the tumour polypeptides and proteins in tumour cells; and
XX CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
XX CC library using well known techniques
XX SQ Sequence 400 BP; 57 A; 156 C; 135 G; 52 T; 0 U; 0 Other;

Query Match 12.8%; Score 41; DB 6; Length 400;
Best Local Similarity 58.7%; Pred. NO. 0.95;
Matches 71; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 90 GCCGCCGACGTTCCGTCGCGGACGACGCGCGGCCCAACGCGACGCGACCTCGAGGC 149
Db 65 GCGGGAAGAGGCGGCTCCACAGTGCAGACCTCGCGCCAGCGCCAGCGCGCTCTCGC 124
Qy 150 TATCGCGGAGCGGTACGCGGAGATGGTGGCGACATCACACGAATTCGACGACACAGC 209
Db 125 CTCGCGGCTTGGCAGGCGGACGACATGCGCGCGTATCGCTCGAAGAGGCGCGGAGCTCATC 184
Qy 210 C 210
Db 185 C 185

RESULT 11
ACA38286
ID ACA38286 standard; DNA; 951 BP.

```

```

XX ACA38286;
XX 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #19943.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Mycobacterium bovis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX FA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX PI WPI; 2003-029926/02.
XX DR P-PSDB; ABU34416.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 26156; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the nucleic acid;
XX (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 951 BP; 141 A; 308 C; 329 G; 173 T; 0 U; 0 Other;

Query Match 12.6%; Score 40.4; DB 8; Length 951;

```

Best Local Similarity 50.4%; Pred. No. 1.4;  
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;  
QY 72 TGGGCGCAACATGTTGGCGCGCGAGCTTCCGTCGCGAGAGCGCGCGCCCAACG 131  
Db 276 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGGTTTACCGACTAACGGTGGT 335  
QY 132 GCAGCGGACCTCGAGGCTATCCCGCGAGCGTA-----CGCGAGATGTTGGCGACATC 185  
Db 336 CAATCGGACGCTCTGGCCCTGCGCGGAGGATCTAGCGCGCGCGCGCGCGGTTGGT 395  
QY 186 ACAGCAATCGACGACGACGACGCGAACTGGCGCTTTTTCGATGATCTCGACGATGA 245  
Db 396 TGCCAAATCTGCGGTACAACTAGCGGTACCGCGCTTCTTGCATCTGCTTGTGAGTTCCC 455  
QY 246 GCAGCGCGGCTTCAGCGGCGGATGAGCTCGGTCGATCCGATATCACTTCCCGACGA 305  
Db 456 GTCGATCCGTGTCGACGTTGATGGTGCAGCGCGAGGTCGCCGAACGGCTCGCGCCGA 515  
QY 306 ACCGACGACAAACAG 321  
Db 516 GCGGCGACGAAAGAG 531

RESULT 12  
AAH51985  
ID AAH51985 standard; DNA; 954 BP.  
XX AC AAH51985;  
DT 04-SEP-2001 (first entry)  
XX Mycobacterium tuberculosis potential drug target gene SEQ ID 39.  
DE Drug target; growth; organism viability; characterisation; ds.  
XX Mycobacterium tuberculosis.  
OS WO200135317-A1.  
XX 17-MAY-2001.  
XX 13-NOV-2000; 2000WO-US031152.  
XX 12-NOV-1999; 99US-0165086P.  
PR 12-NOV-1999; 99US-0165124P.  
PR 01-FEB-2000; 2000US-0179531P.  
XX (REGC ) UNIV CALIFORNIA.  
XX Eisenberg D, Rotstein SH, Marcotte EM;  
PI WPI; 2001-329193/34.  
DR P-PSDB; AAG81134.  
XX Identifying nucleotide or polypeptide sequence for use as drug target,  
PT involves providing algorithm that analyzes a functional relationship  
PT between nucleotide or polypeptide sequences, and comparing the sequences.  
XX Disclosure; Page 76-77; 207pp; English.

This invention relates to a method for identifying a nucleotide or  
XX polypeptide sequence that may be a drug target, or essential for growth  
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
CC tuberculosis proteins which are potential drug targets. The DNA and  
CC protein sequences are used to illustrate the method of the invention. The  
CC method involves providing an unknown nucleotide or polypeptide sequences,  
CC and comparing it to a number of sequences along with at least one  
CC algorithm capable of analysing a functional relationship between  
CC nucleotide and polypeptide sequences. The method is useful for  
CC characterising the function of nucleic acids and polypeptides that may be  
CC useful as a target for a drug or essential for the growth or viability of

CC an organism  
XX SQ Sequence 954 BP; 142 A; 309 C; 330 G; 173 T; 0 U; 0 Other;  
Query Match 12.6%; Score 40.4; DB 4; Length 954;  
Best Local Similarity 50.4%; Pred. No. 1.4;  
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;  
QY 72 TGGGCGGCAACATGTTGGCGCGCGAGCTTCCGTCGCGAGAGCGCGCGCCCAACG 131  
Db 276 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGGTTTACCGACTAACGGTGGT 335  
QY 132 GCAGCGGACCTCGAGGCTATCCCGCGAGCGTA-----CGCGAGATGTTGGCGACATC 185  
Db 336 CAATCGGACGCTCTGGCCCTGCGCGGAGGATCTAGCGCGCGCGCGCGGTTGGT 395  
QY 186 ACAGCAATCGACGACGACGACGCGAACTGGCGCTTTTTCGATGATCTCGACGATGA 245  
Db 396 TGCCAAATCTGCGGTACAACTAGCGGTACCGCGCTTCTTGCATCTGCTTGTGAGTTCCC 455  
QY 246 GCAGCGCGGCTTCAGCGGCGGATGAGCTCGGTCGATCCGATATCACTTCCCGACGA 305  
Db 456 GTCGATCCGTGTCGACGTTGATGGTGCAGCGCGAGGTCGCCGAACGGCTCGCGCCGA 515  
QY 306 ACCGACGACAAACAG 321  
Db 516 GCGGCGACGAAAGAG 531

RESULT 13  
ACA40430  
ID ACA40430 standard; DNA; 954 BP.  
XX AC ACA40430;  
DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #22087.  
DE Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Mycobacterium tuberculosis.  
OS WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU36560.

New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 28300; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:



WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match

Best Local Similarity

Matches 129; Conservative

12.6%; 50.4%;

Score 40.4; DB 4; Pred. No. 2.2;

Length 110000;

OY	72	TGGGGCGCAACATGGTTGGCGGCCCGACGTTCCGTCGCCGAGCAGCGCGGCCCAACG	131
Db	29425	TCGGCTGCAACACAGCCGTGGCGGAGCACTCGCACAGCGAGTTTACCGACTAACGGTGGT	29484
OY	132	GCAGCGGACCTCGAGCTATCGCCGAGCGTA-----CGCCGAGATGGTGGCGACATC	185
Db	29485	CAATCGGAGCTCTGCGCCTGCGCGGAGGATCTAGCCGCGCCGACCGCGTGGT	29544
OY	186	ACACGAAATCGACGACGACACACCGCAACTGGCGCTGTTGTCGATGATCTCGACGATGA	245
Db	29545	TGCCAATCTGCCGTACAACGTAGCGGTACCGCGGTTGTTGCATCTGCTTGTGAGTTCCC	29604
OY	246	GCAGCCCGCGCTTGGCGGGATGAAGTCGGCTGGGATCGGTATCACTTCCCGACGA	305
Db	29605	GTGATCCCGTGTGTCACCGTGATGTGAGGCCGAGGTGCGCGAACGGCTCGCCCGCA	29664
OY	306	ACCCGACAGCAACAG	321
Db	29665	GCCGGCAGCAAGAG	29680

Search completed: February 10, 2005, 05:34:17  
Job time : 346 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:45:06 ; Search time 2015 Seconds  
(without alignments)  
7533.496 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321  
Sequence: 1 ttgaccacaagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	6	AR456084
2	321	100.0	324	6	AX074032 Sequence
3	321	100.0	110000	1	Continuation (30 o
4	321	100.0	346051	1	BX842580 Mycobacte
5	309	96.3	42061	1	AD000019 Mycobacte
6	47.2	14.7	125020	9	AF429315 Homo sapi
7	46.2	14.4	64508	1	AJ564903 Unculture
8	46.2	14.4	67066	1	AB063332 Delftia a
9	46.2	14.4	87888	1	AY365053 Wautersia
10	46.2	14.4	108845	1	U66917 Pseudomonas
11	45.4	14.1	301457	1	AE016924 Chromobac
12	44.8	14.0	320150	1	AP005033 Streptomy
13	44.4	13.8	12201	1	AE012255 Xanthomon
14	44.2	13.8	104289	8	CNS08C7V
15	44.2	13.8	110000	2	LMFLCHR34_10
16	43.8	13.6	13121	1	AE012257 Xanthomon
17	43.6	13.6	303550	1	AL939118 Streptomy
18	43	13.4	2832	8	AY452532 Chlamydom
19	42.8	13.3	13486	1	AE004842 Pseudomon

20	42.8	13.3	71286	1	AF143772	AF143772 Mycobacte
21	42.8	13.3	301068	1	AE017231	AE017231 Mycobacte
22	42.6	13.3	131946	8	AC129226	AC129226 Oryza sat
23	42.4	13.2	646	8	AK062705	AK062705 Oryza sat
24	42.4	13.2	168560	8	AC118132	AC118132 Oryza sat
25	42	13.1	299750	1	AP005964	AP005964 Bradyrhiz
26	42	13.1	299925	1	AP005039	AP005039 Streptomy
27	41.8	13.0	106873	14	AB096160	AB096160 Cercophit
28	41.8	13.0	147891	8	AP004693	AP004693 Oryza sat
29	41.8	13.0	163371	8	AP004585	AP004585 Oryza sat
30	41.6	13.0	10507	1	AE005681	AE005681 Caulobact
31	41.4	12.9	216050	1	AL646076	AL646076 Ralstonia
32	41.2	12.8	289308	1	AE017242	AE017242 Mycobacte
33	41.2	12.8	309267	1	AE017235	AE017235 Mycobacte
34	41	12.8	400	6	CQ462190	CQ462190 Sequence
35	41	12.8	10772	1	AE011659	AE011659 xanthomon
36	40.8	12.7	299050	1	SC0939104	AL939104 Streptomy
37	40.8	12.7	300450	1	AP005960	AP005960 Bradyrhiz
38	40.8	12.7	300917	1	AE016919	AE016919 Chromobac
39	40.8	12.7	349260	1	BX572595	BX572595 Rhodospheu
40	40.6	12.6	4161	8	AK072093	AK072093 Oryza sat
41	40.6	12.6	86896	1	RCU57682	U57682 Rhodobacter
42	40.6	12.6	151269	8	AC016781	AC016781 Genomic S
43	40.6	12.6	160365	2	AC084764	AC084764 Oryza sat
44	40.6	12.6	299425	1	AP005037	AP005037 Streptomy
45	40.6	12.6	314100	1	SC0939106	AL939106 Streptomy

ALIGNMENTS

RESULT 1  
AR456084  
LOCUS AR456084 321 bp DNA linear PAT 20-FEB-2004  
DEFINITION Sequence 93 from patent US 6686166.  
ACCESSION AR456084  
VERSION AR456084.1 GI:42691080  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 321)  
AUTHORS Behr, M., Small, P., Schoolnik, G. and Wilson, M.A.  
TITLE Molecular differences between species of the M. tuberculosis complex  
JOURNAL Patent: US 6686166-A 93 03-FEB-2004;  
FEATURES Location/Qualifiers  
source 1..321  
/organism="unknown"  
/mol\_type="genomic DNA"

Query Match	100.0%	Score 321	DB 6	Length 321
Best Local Similarity	100.0%	Pred. No. 5.9e-49		
Matches 321	Conservative	0	Mismatches	0
Indels	0	Gaps	0	
Qy	1	TTGACCCCAAGCGGCACTAAACGCCAGCGGCTCGCCAGGCTCAACGCCCGCGT	60	
Db	1	TTGACCCCAAGCGGCACTAAACGCCAGCGGCTCGCCAGGCTCAACGCCCGCGT	60	
Qy	61	CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGCGGCTCGCCAGGCTCAACGCCCGCGT	120	
Db	61	CGGAATCGGTTGGCGGCGGCAACATGTTGGCGGCGGCGGCTCGCCAGGCTCAACGCCCGCGT	120	
Qy	121	CGCGCCCAAGCGGCGGCACTCGAGGCTATCGCCGCGGCTAGCCGAGATGTTGGCG	180	
Db	121	CGCGCCCAAGCGGCGGCACTCGAGGCTATCGCCGCGGCTAGCCGAGATGTTGGCG	180	
Qy	181	ACATCACAGGAATCGAGCGACACAGCCGCACTGGGGCTGTTGTCGATGATCTTCGAC	240	
Db	181	ACATCACAGGAATCGAGCGACACAGCCGCACTGGGGCTGTTGTCGATGATCTTCGAC	240	
Qy	241	GATGACGACGCGCGGCTTTGAGCGGGGATGAAGTCGGCTGGCATCCGTCATCATTCCCC	300	

Db 241 GATGAGCAGCGCGCTTGAGCGGGGATGAAGTCGGTGGCATCGTATCACTTCCCC 300  
QY 301 GACGAACCCGACGAAACAG 321  
Db 301 GACGAACCCGACGAAACAG 321

## RESULT 2

AX074032  
LOCUS AX074032 324 bp DNA linear PAT 06-FEB-2001  
DEFINITION Sequence 14 from Patent WO0104151.  
ACCESSION AX074032  
VERSION AX074032.1 GI:12710257  
KEYWORDS  
SOURCE  
ORGANISM

Mycobacterium tuberculosis  
Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

## REFERENCE

1 Andersen, P. and Skjoet, R.  
Tuberculosis vaccine and diagnostics based on the mycobacterium tu  
berculosis eat-6 gene family  
Patent: WO 0104151-A 14 18-JAN-2001;  
Statens Serum Institut (DK)

## FEATURES

source  
1..324  
Location/Qualifiers  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"

## CDS

1..324  
/notes="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC28442.1"  
/db\_xref="GI:12710258"  
/translation="MTHKTKRQPAIAAGINAPRRNRVGRHQGWADVPASBQRAQR  
QRDLAIRRAYAEWATVSHSIDDDTAELALLSHLDDQRLEAGMKLQWHPYHFPDE  
PDSKQ"

## ORIGIN

Query Match 100.0%; Score 321; DB 6; Length 324;  
Best Local Similarity 100.0%; Pred. No. 5.9e-49;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTGACCCACAAGCGCACTAAACGCCAGCGCATCGCGCGGCTCAACGCCCGCGT 60  
Db 1 TTGACCCACAAGCGCACTAAACGCCAGCGCATCGCGCGGCTCAACGCCCGCGT 60

QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGCGGCGAGCTTCGGTCCGCCGAGCAGCGC 120  
Db 61 CGGAATCGGTTGGCGGCAACATGTTGGCGCGGCGAGCTTCGGTCCGCCGAGCAGCGC 120

QY 121 CGCGCCCAACGCGGCACTCGAGGTATCCGCCGAGCGTACGCCGAGATGTTGGCG 180  
Db 121 CGCGCCCAACGCGGCGCACTCGAGGTATCCGCCGAGCGTACGCCGAGATGTTGGCG 180

QY 181 ACATCACAGAAATCGACGACGACGCGGCTGTTGTCGATGATCTTCGAC 240  
Db 181 ACATCACAGAAATCGACGACGACGCGGCTGTTGTCGATGATCTTCGAC 240

QY 241 GATGAGCAGCGCGGTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300  
Db 241 GATGAGCAGCGCGGTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300

QY 301 GACGAACCCGACGAAACAG 321

Db 301 GACGAACCCGACGAAACAG 321

## RESULT 3

AE000516\_29/c

## WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name	Begin	End
AE000516_00	1	110000
AE000516_01	100001	210000
AE000516_02	200001	310000
AE000516_03	300001	410000
AE000516_04	400001	510000
AE000516_05	500001	610000
AE000516_06	600001	710000
AE000516_07	700001	810000
AE000516_08	800001	910000
AE000516_09	900001	1010000
AE000516_10	1000001	1110000
AE000516_11	1100001	1210000
AE000516_12	1200001	1310000
AE000516_13	1300001	1410000
AE000516_14	1400001	1510000
AE000516_15	1500001	1610000
AE000516_16	1600001	1710000
AE000516_17	1700001	1810000
AE000516_18	1800001	1910000
AE000516_19	1900001	2010000
AE000516_20	2000001	2110000
AE000516_21	2100001	2210000
AE000516_22	2200001	2310000
AE000516_23	2300001	2410000
AE000516_24	2400001	2510000
AE000516_25	2500001	2610000
AE000516_26	2600001	2710000
AE000516_27	2700001	2810000
AE000516_28	2800001	2910000
AE000516_29	2900001	3010000
AE000516_30	3000001	3110000
AE000516_31	3100001	3210000
AE000516_32	3200001	3310000
AE000516_33	3300001	3410000
AE000516_34	3400001	3510000
AE000516_35	3500001	3610000
AE000516_36	3600001	3710000
AE000516_37	3700001	3810000
AE000516_38	3800001	3910000
AE000516_39	3900001	4010000
AE000516_40	4000001	4110000
AE000516_41	4100001	4210000
AE000516_42	4200001	4310000
AE000516_43	4300001	4403837

Continuation (30 of 44) of AE000516 from base 2900001 (AE000516 Mycobacterium tuberculosis)

Query Match	Best Local Similarity	Score 321; DB 1; Length 110000;	Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGACCCACAAGCGCACTAAACGCCAGCGCATCGCGCGGCTCAACGCCCGCGT 60	100.0%;	Score 321; DB 1; Length 110000;	Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 71695 TTGACCCACAAGCGCACTAAACGCCAGCGCATCGCGCGGCTCAACGCCCGCGT 71636			
QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGAGTTCGGTCCGCCGAGCAGCGC 120			
Db 71635 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGGAGTTCGGTCCGCCGAGCAGCGC 71576			
QY 121 CGCGCCCAACGCGGCACTCGAGGTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 180			
Db 71575 CGCGCCCAACGCGGCGCACTCGAGGTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 71516			
QY 181 ACATCACAGAAATCGACGACGACGCGGCTGTTGTCGATGATCTTCGAC 240			
Db 71515 ACATCACAGAAATCGACGACGACGCGGCTGTTGTCGATGATCTTCGAC 71456			
QY 241 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300			
Db 71455 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 71396			

QY 301 GACGAACCGACAGCAACAG 321  
 Db 71395 GACGAACCGACAGCAACAG 71375

RESULT 4  
 LOCUS BX842580/c  
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 9/13.  
 ACCESSION BX842580 AL008967 AL021185 AL021186 AL021246 AL123456  
 277250 277724 280225 283863 295387 295556 296072 297051 298045  
 298209

VERSION BX842580.1 GI:41353422  
 KEYWORDS complete genome.  
 SOURCE Mycobacterium tuberculosis H37Rv  
 ORGANISM Mycobacterium tuberculosis H37Rv  
 Bacteria; Actinobacteria; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.

REFERENCE 1  
 AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eiglmeier, K., Gao, S., Barry, III, C.E.,  
 Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
 Barrall, B.G.  
 Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence  
 NATURE 393 (6685), 537-544 (1998)

REFERENCE 2  
 AUTHORS Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.  
 TITLE Re-annotation of the genome sequence of Mycobacterium tuberculosis  
 H37Rv  
 JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)  
 MEDLINE 22255591  
 PUBMED 12368430

REFERENCE 3 (bases 1 to 346051)  
 AUTHORS Parkhill, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On or before Jan 28, 2004 this sequence version replaced  
 gi:3261488, gi:3261504, gi:3261505, gi:3261507, gi:3261617,  
 gi:3261620, gi:3242265, gi:3261685, gi:3261763, gi:3261778,  
 gi:3261793, gi:3261814, gi:3261840, gi:3261838.

NOTES:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).

FEATURES  
 source  
 1..346051  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 142..528  
 /gene="glbO"  
 /locus\_tag="RV2470"  
 /function="OXYGEN CARRIER, INVOLVED IN OXYGEN TRANSPORT."  
 /note="RV2470, (MTV008.26), len: 128 aa. Possible glbO,  
 globin-like protein, highly similar to Q9CC59[GLB0]ML1253  
 HEMOGLOBIN-LIKE (OXYGEN CARRIER) from Mycobacterium leprae

gene  
 142..528  
 /gene="glbO"  
 /locus\_tag="RV2470"  
 /function="OXYGEN CARRIER, INVOLVED IN OXYGEN TRANSPORT."  
 /note="RV2470, (MTV008.26), len: 128 aa. Possible glbO,  
 globin-like protein, highly similar to Q9CC59[GLB0]ML1253  
 HEMOGLOBIN-LIKE (OXYGEN CARRIER) from Mycobacterium leprae

CDS  
 142..528  
 /locus\_tag="RV2470"  
 /function="OXYGEN CARRIER, INVOLVED IN OXYGEN TRANSPORT."  
 /note="RV2470, (MTV008.26), len: 128 aa. Possible glbO,  
 globin-like protein, highly similar to Q9CC59[GLB0]ML1253  
 HEMOGLOBIN-LIKE (OXYGEN CARRIER) from Mycobacterium leprae

(128 aa), FASTA scores: opt: 767, E(): 4e-47, (88.1%  
 identity in 126 aa overlap); O9X7B3[MLCB1610.14C PUTATIVE  
 GLOBIN from Mycobacterium leprae (131 aa);  
 O9L250[SC6D10.14 PUTATIVE GLOBIN from Streptomyces  
 coelicolor (137 aa), FASTA scores: opt: 466, E(): 5.7e-26,  
 (53.6% identity in 125 aa overlap). Also similar to O31607  
 YJBI PROTEIN from Bacillus subtilis (132 aa), FASTA  
 scores: opt: 294, E(): 6.6e-14; (39.85% identity in 128 aa  
 overlap). COULD BELONG TO PROTOZOAN/CYANOBACTERIAL GLOBIN  
 FAMILY PROTEIN."  
 /codon\_start=1  
 /transl\_table=11  
 /product="POSSIBLE GLOBIN (OXYGEN-BINDING PROTEIN) GLBO"  
 /protein\_id="CAA16047.1"  
 /db\_xref="GI:2791510"  
 /translation="MPKSFYDAVGAKTDAIVSRFYAQAQAEDEVLRVYPEDDLAGA  
 EERLRMLEQYWGSPRTYSEQRGPHLRMRHAPRISLIERDAWLRCHMTAVASIDSE  
 TLDEHREKLQYDLEMAHSLVNSPF"  
 528..2168  
 /gene="agla"  
 /locus\_tag="RV2471"  
 528..2168  
 /gene="agla"  
 /locus\_tag="RV2471"  
 /EC\_number="3.2.1.20"  
 /function="INVOLVED IN SUGAR METABOLISM (HYDROLYSIS OF  
 TERMINAL, NON-REDUCING 1,4-LINKED D-GLUCOSE RESIDUES WITH  
 RELEASE OF D-GLUCOSE)."  
 /note="RV2471, (MTV008.27), len: 546 aa. Probable agla,  
 maltase (alpha-glucosidase) (EC 3.2.1.20), highly similar  
 or similar to several e.g. Q60027[AGLA from  
 Thermomonospora curvata (544 aa), FASTA scores: opt: 2071,  
 E(): 4e-116, (57.7% identity in 525 aa overlap);  
 O9K2E3[AGLAE from Streptomyces coelicolor (534 aa), FASTA  
 scores: opt: 1475, E(): 1.5e-80, (50.1% identity in 537 aa  
 overlap); O86874[AGLA from Streptomyces lividans (534 aa),  
 FASTA scores: opt: 1473, E(): 2e-80, (50.1% identity in  
 537 aa overlap); etc. SEEMS TO BELONG TO FAMILY 13 OF  
 GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE  
 FAMILY."  
 /codon\_start=1  
 /transl\_table=11  
 /product="PROBABLE ALPHA-GLUCOSIDASE AGLA (MALTASE)  
 (GLUCOSINVERTASE) (GLUCOSIDOSUCRASE) (MALTASE-GLUCOAMYLASE)  
 (LYSOSOMAL ALPHA-GLUCOSIDASE) (ACID MALTASE)"  
 /protein\_id="CAA16048.1"  
 /db\_xref="GI:2791511"  
 /translation="MDQQRDPDMPGSPRASARRPEPDPMPGEPWRSRAVYQVYPRS  
 FADSDGVDGLDGLASLKLHLQGLVDALINPVTVSPMDHGDYDADPRDIDPLFG  
 GMPAFERLVAARHQGIKVTDVVPNHTSSAHFQALADLPGPSARDRYFFRDGRG  
 PDGSLPPNMWESVFGPAWTRPDPGNGPWYHLHFDTEQDPLNDNPEILDDFEKT  
 LRFWLDGVDGFRIDVAHMAKPGCLPSPDLGIEVLHRRDDPRPNHNVHAIHRDI  
 RTVIDPGATVGEVWHDNAWAEVLRPDHLHGFRLARFEFDAEIRDAVANS  
 LAAALQCATPTILANHDVGRSRYGGEGIGLRKAKAMAVMLALPGVFLYNGQE  
 LGLPVDLDPDVLQSTPSSRTERGDRCPVLPWSGNIPFPFGFTCPDPTWLMPP  
 EWAALTAEKQADAGTSIFRLALRLRERNEFDGVDVLAAPDADALIFRRHGGGLV  
 CALNAERPLALPAGEPILASAPLTADLTLPNAAALV"  
 2236..2529  
 /locus\_tag="RV2472"  
 2236..2529  
 /locus\_tag="RV2472"  
 /function="UNKNOWN"  
 /note="RV2472, (MTV008.28), len: 97 aa. Conserved  
 hypothetical protein, showing some similarity to  
 O53451[Rv1103c]MTV017.56c from Mycobacterium tuberculosis  
 strain H37Rv (106 aa), FASTA scores: opt: 135, E(): 0.026,  
 (45.85% identity in 72 aa overlap); and AAK45393[MT1135  
 HYPOTHETICAL 11.4 KDA PROTEIN from Mycobacterium  
 tuberculosis strain CDC1551 (78 aa) FASTA scores: opt:  
 139, E(): 0.011, (45.35% identity in 75 aa overlap)."  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11

```
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAAL6049.1"
/db_xref="GI:2791512"
/translation="MMKRIAVRLPGEVITFDSEVSOIRPSRAAVVLRASNASDAA
ILTATEPNHLDLALGQAALPTSIDAAHPARPARRDPCLYPRTGQALPRTG"
gene
2532..3248
CDS
2532..3248
/locus_tag="RV2473"
/locus_tag="RV2473"
/functions="UNKNOWN"
/notes="RV2473, (MTV008.29), len: 238 aa. Possible
pro-,ala-rich membrane protein, with possible
transmembrane domain around aa 81-104."
/codon_start=1
/transl_table=11
/product="POSSIBLE ALANINE AND PROLINE RICH MEMBRANE
PROTEIN"
/protein_id="CAAL6050.1"
/db_xref="GI:2791513"
/translations="MAPTSSVASSELLMPFSAAGVGVWRTTATASORYHRPMSDT
PRAEPYEPORPPGVPDGDGSRPAPSTTPSPGRWSLVAALAVSLGVIGVW
FHRQPDHSPASATFTSQISDAKENVCAAHRIVRQAVLNTQANPVPDPTGD
LAVANARALYSGGYLLRLRTAEPATPAELRDAVRSLANALQELAVNYLAGPDSV
VPLRLALERTDRAVDPLCV"
complement(3280..3933)
/locus_tag="RV2474c"
complement(3280..3933)
/locus_tag="RV2474c"
/functions="UNKNOWN"
/notes="RV2474c, (MTV008.30c), len: 217 aa. Hypothetical
protein. Shows weak similarity with Q9L246|SCSD10.18c
HYPOTHETICAL 24.9 KDA PROTEIN from Streptomyces coelicolor
(238 aa). FASTA scores: opt: 111, E(): 5.6, (30% identity
in 233 aa overlap), BLASTP scores: Score= 135, E= 3.5e-07,
P= 3.5e-07, Identities= 55/182 (30%)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAAL6051.1"
/db_xref="GI:2791514"
/translations="WVERGLWLPDPAHRADLAFVDFHALLRDLDDAVIRARSTGLLS
AWATGFDVLASRVAGKVRPDDLVAARSALHGLATTDSAGYDFVGYSDMSAWRGL
PPESGTYLDDVDFARVWMLDRAHRAKEHSSAGPVPVSLDQEVIVQVSSADVVVGL
PMECVFALTAMGFLPQSAETISADELIRVISPALRLDARFSGVYVHRGHAALVLR"
complement(3939..4355)
/locus_tag="RV2475c"
complement(3939..4355)
/locus_tag="RV2475c"
/functions="UNKNOWN"
/notes="RV2475c, (MTV008.31c), len: 138 aa. Conserved
hypothetical protein, showing similarity with
Q9L245|SC6D10.19c HYPOTHETICAL 16.2 KDA PROTEIN from
Streptomyces coelicolor (136 aa). FASTA scores: opt: 236,
E(): 1.9e-09, (34.1% identity in 126 aa overlap). Also
some similarity with AKA44393|Z97050|MTCI28.3 CONSERVED
HYPOTHETICAL PROTEIN from Mycobacterium tuberculosis
cosmid I (151 aa). FASTA scores: opt: 147, E(): 0.00025,
(29.2% identity in 120 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/Query Match
Best Local Similarity 100.0%; Score 321; DB 1; Length 346051;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGACCCACAAAGCGCACTAAACCCACGAGCCATCGCGCGAGGCTCAACGCCCGCGGT 60
Db 203873 TTGACCCACAAAGCGCACTAAACCCACGAGCCATCGCGCGAGGCTCAACGCCCGCGGT 203814
QY 61 CGGAATCGCGTTCGGCGGCAACATGTTGGCGGCGAGCTTCGTCGCGAGAGCGGC 120
Db 203813 CGGAATCGCGTTCGGCGGCAACATGTTGGCGGCGAGCTTCGTCGCGAGAGCGGC 203754
```

```
QY 121 CGGCCCCAACGCGAGCGGACCTCGAGGCTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 180
Db 203753 CGGCCCCAACGCGAGCGGACCTCGAGGCTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 203694
QY 181 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
Db 203693 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 203634
QY 241 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTCATCACTTTCCCC 300
Db 203633 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTCATCACTTTCCCC 203574
QY 301 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 321
Db 203573 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 203553

RESULT 5
MSGY223
LOCUS MSGY223 42061 bp DNA linear BCT 10-DEC-1996
DEFINITION Mycobacterium tuberculosis sequence from clone y223.
ACCESSION AD000019
VERSION AD000019.1 GI:1717737
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 42061)
AUTHORS Du,L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT GSDB:S:1004712.
FEATURES
source
1..42061
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y223"
ORIGIN
Query Match 96.3%; Score 309; DB 1; Length 42061;
Best Local Similarity 99.7%; Pred. No. 4e-47; Mismatches 1; Gaps 1;
Matches 320; Conservative 0; Indels 0;
QY 1 TTGACCCACAAAGCGCACTAAACCCACGAGCCATCGCGCGAGGCTCAACGCCCGCGGT 60
Db 5783 TTGACCCACAAAGCGCACTAAACCCACGAGCCATCGCGCGAGGCTCAACGCCCGCGGT 5842
QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGAGCTTCGTCGCGCGAGAGCGGC 120
Db 5843 CGGAATCGGTTGGCGGCAACATGTTGGCGGCGAGCTTCGTCGCGCGAGAGCGGC 5902
QY 121 CGGCCCCAACGCGAGCGGACCTCGAGGCTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 180
Db 5903 CGGCCCCAACGCGAGCGGACCTCGAGGCTATCCGCCGAGCGGTACGCCGAGATGTTGGCG 5962
QY 181 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
Db 5963 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6022
QY 241 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTCATCACTTTCCCC 300
Db 6023 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTCATCACTTTCCCC 203574
QY 301 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 321
Db 6082 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6102
```

```
RESULT 6
AF429315/c      125020 bp      DNA      linear      PRI 18-JAN-2002
LOCUS      Homo sapiens juncctophilin 3 (JPH3) gene, partial cds.
DEFINITION
ACCESSION      AF429315
VERSION      AF429315.1 GI:17646244
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding juncctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
11694876
REFERENCE
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11694876
REFERENCE
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11694876
FEATURES
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
Disease-Like 2 (HDL2)."
complement(35561..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(<36507..>36887)
/gene="JPH3"
/note="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="juncctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="juncctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGGRNFDDGSGYCGWEDGKAHGVCVTPKPGCGEYTCGWS
HGFEVIGYVTPSGNVTQGTWAGQRRHGIGLESKWKYKGWTHGFKRGYGVRECA
NGAKYEGTWSNGLODQGYGTETYSVG"
ORIGIN
Query Match      14.7%; Score 47.2; DB 9; Length 125020;
Best Local Similarity 10.6%; Pred. No.19;
Matches 31; Conservative 149; Mismatches 112; Indels 0; Gaps 0;
QY      20 AACGCAGCAGCCATCGCCGAGGCTCAACGCCCGCGTCCGAATCGGTGGCGGC 79
Db      51134 MACMYWYKSRRTKSMATSGCWRMGAMRSKGMRYKWSCRYKGMRWGWSYKCSGYM 51075
QY      80 AACATGTTGGCGCGCGAGTTCGTCCGCGAGCAGCGCGCCCAACGCGAGCGG 139
Db      51074 SMGMTGSGRTSYCYTGGYCYCKSAKMKCKSKTSKSCYKSGMRYSYCTCWGG 51015
QY      140 ACCTGAGCGCTATCGCCGAGGTAGCGCGAGATGTCGCGACATCACAGAAATCGAG 199
Db      51014 RMCKGMKSSCASRSCKSKYKAGWSRCKRGRMGSCWYWSWKSYMSASYKSASRGMK 50955
QY      200 ACGACACAGCCGAACTGGCGCTGTTGTCATGATCATCTCGACGATGAGCAGCCGCGCTTG 259
```

```
Db      50954 CYKTSCKMYYSMBKSHDBSGYKYSKYBYWTMMKDSMCVSDSHMMRDMYDKMR 50895
QY      260 AGGCGGGGATGAAGCTCGCTGGCATCGGTATCACTTCCCGCAGCAACCCGA 311
Db      50894 MKVSSGTRGHBHSAWCKMSRSDSGYSGMYHMKSYWKSASKHMYYSGM 50843

RESULT 7
UBA564903/c      64508 bp      DNA      circular BCT 22-JAN-2004
LOCUS      Uncultured bacterium Incp-ibeta multiresistance plasmid pB10.
ACCESSION      AJ564903
VERSION      AJ564903.1 GI:31746361
KEYWORDS      'tnpA gene; acetyltransferase;
aminoglycoside-3'-phosphotransferase;
aminoglycoside-6-phosphotransferase; ATPase, autophosphorylase;
beta-lactamase precursor; coregulator protein; damage
inducible-like protein; dihydropteroate synthetase; din gene; DNA
integrase; DNA relaxase; DNA replication primase; DNA
topoisomerase; DNA transport protein; hypothetical transmembrane
protein; hypothetical Xf2080 protein; incC1 gene; incC2 gene; incC2
protein; inclusion membrane protein; integral membrane protein for
mercuric transport; int11 gene; kfrA gene; KfrA protein; klcA gene;
KlcA protein; klcB gene; KlcB protein; klcE gene; KlcE protein;
KlcE gene; KlcE protein; klcF gene; KlcF protein; klcA gene; KorA
protein; korB gene; KorB protein; korC gene; korC protein;
maturation peptidase; merA gene; mercuric reductase; mercury
resistance protein; merD gene; merE gene; merP gene; merR gene;
merT gene; muramidase; ORF-2; ORF5; ORFE-like; origin of transfer;
oriT; oriT binding protein; outer membrane protein; oxa-2 gene;
peCM-like gene; periplasmic mercury ion binding protein; prepilin;
qacEdelta1 gene; relaxosome stabilisation protein; replication
initiation protein; repressor protein; resolvase; single-strand DNA
binding protein; small exporter protein; sbg gene; strA gene; strB
gene; sul1 gene; teta gene; tetr gene; tetracycline efflux protein;
tetracycline resistance repressor protein; tnpA gene; tnpA' gene;
tnpR gene; trac gene; traB gene; traI gene; traJ gene; traK gene; traL
gene; traG gene; traH gene; traI gene; traJ gene; traK gene; traL
gene; traM gene; traM protein; traN gene; traN protein; transfer
origin protein; trac gene; traO protein; trbA gene; trbA protein;
trbS gene; trbc gene; trbd gene; trbd protein; trbS gene; trbE
protein; trbF gene; trbF protein; trbg gene; trbg protein; trbH
gene; trbI gene; trbi gene; trbi protein; trbj gene; trbj
protein; trbk gene; trbk protein; trbl gene; trbm gene; trbm
protein; trbn gene; trbo gene; trbo protein; trbp gene; trbp
protein; trfai gene; truncated transposase; upf30.5 gene; xif2080
gene.
SOURCE      uncultured bacterium
ORGANISM      uncultured bacterium
Bacteria; environmental samples.
REFERENCE
1 Schluter,A., Heuer,H., Szczepanowski,R., Forney,L.J., Thomas,C.M.,
Puhler,A. and Top,E.M.
The 64 508 bp IncP-ibeta antibiotic multiresistance plasmid pB10
isolated from a waste-water treatment plant provides evidence for
recombination between members of different branches of the
Incp-ibeta group
Microbiology (Reading, Engl.) 149 (Pt 11), 3139-3153 (2003)
22962753
JOURNAL      MEDLINE
PUBMED      14600226
REFERENCE
2 (bases 1 to 64508)
Schlueter,A.
Direct Submission
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
14600226
FEATURES
source
1..64508
Location/Qualifiers
/organism="uncultured bacterium"
/mol_type="genomic DNA"
/isolation_source="activated sludge"
/db_xref="taxon:77133"
```

```

/environmental_sample
/lab_host="Pseudomonas sp. B13"
/plasmid="pB10"
97. .459
/genes="trbA"
/genes="trbA"
/function="regulatory protein for mating pair formation"
/codon_start=1
/transl_table=11
/product="trbA protein"
/protein_id="CAD97487.1"
/db_xref="GI:31746362"
/db_xref="TrEMBL:Q7X3F3"
/translation="MNYIPFTNVRLRDLDERHMTKKELSDRSQVSIISFLSLDTTKAN
PSLKMEDIAQALETPLPLELLESTDLKDALDLAGGKAPRSIPPGFERVAALVPEHQ
AFIVKNGEATRKLRGS"
769. .1731
/genes="trbB"
769. .1731
/genes="trbB"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="putative ATPase, autophosphorylase"
/protein_id="CAD97488.1"
/db_xref="GI:31746363"
/db_xref="TrEMBL:Q7X3F2"
/translation="MTDQPEHNSIKERAKKKLERDMGPELLAALNDPKTIVMLNAD
GKMLERLEGPMCKITGLRVAQAIIETIAGHGEVTRSPILGEGLPDGSRFAG
QLPVPVPAFTFAIKVAIFTLDDQYVERGIPTAPQREALIAAFAHNNILVIGTGS
KTTLVNAINEMVIQDPTVERFIIEDTGEIOCAENYVQYHTSIDNNMTALKITLR
MRPRLIVGEVGRPEALDILMAWNTQHEGGAATHANNAKAGLDRLAMLISMHPDSPK
PIRPLIGEAVHVVHIAIVEGSRRIQEIILEVSGFANGYITKTL"
1748. .2212
/genes="trbC"
1748. .2212
/genes="trbC"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="putative prepilin"
/protein_id="CAD97489.1"
/db_xref="GI:31746364"
/db_xref="TrEMBL:Q7X3F1"
/translation="MQATFPAPRLNRNALFYMLGFALLAPFLLAPOHAFASGTTGSL
PYESWTLNRNSVTGPVAFALSTIGIVVAGGILIFGGEIENGFFRTLIFVLWALLVG
AONMSTFFSGRAEIAALTDGALHQVKVVALDVAGPPGEAVMWAERGHAG"
2216. .2527
/genes="trbD"
2216. .2527
/genes="trbD"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbD protein"
/protein_id="CAD97490.1"
/db_xref="GI:31746365"
/db_xref="TrEMBL:Q7X3F0"
/translation="MALRAIPIRAGNRNENLFWGGDRELVMFSGLLAGALIFSAQEV
ATVFGALWFGALFALRLMAKADPKURHVILRHRRYKPYIPARSTPFRDNTPSQKQY
K"
2524. .5082
/genes="trbE"
2524. .5082
/genes="trbE"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbE protein"
/protein_id="CAD97491.1"
/db_xref="GI:31746366"
/db_xref="TrEMBL:Q7X3E9"

```

```

/translation="MIEAIAIAIVGLAVLLILFARAVDAELKLLKHKRAKDAGLA
DLINVAAYVDGVIKKGSSFWAAMLYKGDNDASSTEEOREMVSFRINOAALAGLNGW
MYHVDARPPAPNYSERGVSFDSVSAIDERRRLRGLGAMEGYVLTWTWPPF
VLAQKPFVLMFDDDAVADPHARTITGLIAQKREITLSLSLSSAVGKVRURGQKVV
SEEGREVTDDFLSLQFCVGLGHHVPLPSNPMYLDVAVIGGOELWGGVVPKIGKFI
QVVAJEGFPLESTPGVLSALAEIPSEYRWSSRFIPMDQHESLKHLDKFKKWKQKIRG
PFDQVNTNTGSIINODAAWVGDAEAAIAEVNSGLVAAGYTSVVLMDEDRERLAAS
ALLVKAVNRLAFAARIETINTLDVAGSLGPHGVENVRPLINTMNIALDLPSTSIW
TGSATAPCPMPYPLPALMHCVTVGATPRNLNHRVDLGHTEFMTGPTGAGKSTHLGI
AAQLRRYAGMSIYAFDKCMSYPLAAGIRATKSGKSLHFTVAADDDRLAFGLPQLE
TKGDRAMAWEMIDTILALGVNVTTPAQNEIGNAIMSHASGATLSEFSVPTQDEAI
REAIKQYTVDSGMLHLDDEGLSDFTVEIIEMLNGLKGFALPVLILYLFRIER
ALKQGPSVIIIDEAWMLCHPAFRAKIRWLKRLKANCLVMAQSLDAANSGLID
VIVGETATKIFLPNVVARDDETSALYRRMGLNAROIEILATAI PKROYVYVSENGRRL
YDLALGPMALATVGSADKESVATIKSLEAKFGHEVHWEWLAGRLNLDYLEA"
5079. .5861
/genes="trbF"
5079. .5861
/genes="trbF"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbF protein"
/protein_id="CAD97492.1"
/db_xref="GI:31746367"
/db_xref="TrEMBL:Q7X3E8"
/translation="MSIADTLKGLIFKKPANSAPRDRSRKEPLAGGRAGEENPY
LAARTWHDVGSVVSQRTQVQVIGILSLILALAGVGVHIGSQSKFPIYVVEVDKL
TAAAGPQVQAGKADPRVIHAADVMSKARMSPDVALQRKCVKFKAYSMLAPNDPA
TFKMEWNLGNDSPSPFRAEKEMVSVIEKTVIPQTPDTQVQWVETTRDROGTLKQ
PFWALVTVTTIAEVTNTDQLRNNPLSIYVSDYSWRIQ"
5858. .6778
/genes="trbG"
5858. .6778
/genes="trbG"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbG protein"
/protein_id="CAD97493.1"
/db_xref="GI:31746368"
/db_xref="TrEMBL:Q7X3E7"
/translation="MNSRSSTMKTLTSLIVGAAVAVPSLALAAPGDDIADKYFNKTN
PTLTQAERAAIAAKRWAGSAGATIGKPVAGSGAIFVYGAQQPIVCAVLQVCDVAL
QPEOVNINIGLDTARTWTEPAITGSGATEVQHLLIKPMQVGLTSLVTTNRRSVHF
RLRSRTYMPQVSPFTYPEAQAKWDATQREKQERTDNTLTPATGEVLGDLSDYELS
GSASKPVRVYNDGKRTIIEPMTQOOTEAPTLTVRKEGGLFTDDETFVLNRYVQGN
RIIVDTVFDRAILLIAGVSSQDRVTISRRK"
6781. .7269
/genes="trbH"
6781. .7269
/genes="trbH"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbH protein"
/protein_id="CAD97494.1"
/db_xref="GI:31746369"
/db_xref="TrEMBL:Q7X3E6"
/translation="MRKIVSLALLALGLGCGCATTSQYGNFVQSAALDQOKLATDAV
QQLATLYAPARTLRLEQQPTPDPFGQALVKSRLDKGYALLLEYAPAGASTQAPASASQ
ASAPATTAAAGGLPLRYVLDQAGDSNLYRLTLMVGNQSIITRPPYLAQNGTFAPAGYVVR
KE"
7274. .8695

```

```

Query Match      14.4%; Score 46.2; DB 1; Length 64508;
Best Local Similarity 47.7%; Pred. No. 33;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy      6 CCACAAGCGCACTAAACGCCAGCCAGCCATCGCCAGGGCTCAACAGCCCCCGCGCGGAA 65
Db      35080 CGACAAAGCCCGAGGAGAGGGCCACGCCAGCCAGCGCTCGGCGGCGGA 35021

```

Qy 66 TCAGCTGGCGGCAACATGTTGGCGCGCGAGCGTTCCTCGCTCGCGGAGCGAGCGCGCGC 125  
Db 35020 CCTCGCGCGCGCAACAGCGGTACCGCGAGATCGAGCGCGCGCGCGCGAGCTGCGTGC 34961  
Qy 126 CCAAGCGGAGCGGACCTCGAGGCTATCCGCGGAGCGTACCGCGAGATGTTGGCGACATC 185  
Db 34960 CGACTCGAGCGCGCGAACCAGGCGCGACCGAGCGCGCGCGCGCTCAGCGCGGAGCGCG 34901  
Qy 186 ACAGGAATCCAGCAGCAGCAGCGGAACTCGCGCTGTTGCGATGCATCTCGACGATGA 245  
Db 34900 GAAGCCACCGAGCGCGCGCGCGGAGCGGACAGGTGCGCGCGAGCTGGTGAAGGT 34841  
Qy 246 GCAGCGCGCGCTTGAAGCGCGGATGAAGCTCGCTGGCGATCCG 288  
Db 34840 GCAGCGCAAGCGCGGCGCGGAGCGGCGCGGAGCGGCGCGGAGCGAGCAG 34798

RESULT 8  
AB063332/c 67066 bp DNA circular BCT 16-JUL-2004  
LOCUS AB063332  
DEFINITION Delftia acidovorans plasmid puO1 DNA, complete sequence.  
ACCESSION AB063332  
VERSION AB063332.1 GI:34013299  
KEYWORDS  
SOURCE Delftia acidovorans  
ORGANISM Delftia acidovorans  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Comamonadaceae; Delftia.

REFERENCE  
1 Sota, M., Endo, M., Nitta, K., Kawasaki, H. and Tsuda, M.  
AUTHORS Characterization of a class II defective transposon carrying two  
TITLE haloacetate dehalogenase genes from Delftia acidovorans plasmid  
puO1  
JOURNAL Appl. Environ. Microbiol. 68 (5), 2307-2315 (2002)  
MEDLINE 21972280  
PUBMED 11976102

REFERENCE  
2 Sota, M., Kawasaki, H. and Tsuda, M.  
AUTHORS Structure of haloacetate-catabolic Incp-1beta plasmid puO1 and  
TITLE genetic mobility of its residing haloacetate-catabolic transposon  
J. Bacteriol. 185 (22), 6741-6745 (2003)  
MEDLINE 22957203  
PUBMED 14594853

REFERENCE  
3 (bases 1 to 67066)  
AUTHORS Sota, M., Tsuda, M. and Kawasaki, H.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-2001) Masahiro Sota, Institute for Environmental  
Sciences, Department of Environmental Simulation; 1-7 Ienomae,  
Obuchi, Rokkasho, Aomori 039-3212, Japan (E-mail:sota@ies.or.jp,  
Tel:81-175-71-1407, Fax:81-175-71-1401)  
PEATURES Location/Qualifiers  
1. 67066  
/organism="Delftia acidovorans"  
/mol\_type="genomic DNA"  
/strain="B"  
/db\_xref="taxon:80866"  
/plasmid="puO1"  
/country="Japan:Osaka"  
/note="Incp-beta plasmid-the organism was originally  
submitted as Moraxella sp. B"  
1. .901  
/note="vegetative origin (oriV)"  
897. .901  
/note="target sequence of TnHad2"  
902. .18505  
/note="Tn3-family, Tn21-group, Tn21-subgroup"  
/transposon="TnHad2"  
902. .939  
/rpt\_type="inverted"  
949. .9884  
/transposon="TnHad1"  
949. .4152  
/note="IS1071"

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

repeat\_region  
gene  
CDS

1094. .4009  
/gene="tnpA"  
1094. .4009  
/gene="tnpA"  
/function="transposition"  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="transposase"  
/protein\_id="BAC81973.1"  
/db\_xref="GI:34013300"  
/translation="MQSHHTTFLGMRGLPRDISDFENKAFTEFGAERDAINARRGDS  
HKLGLALHIGFLRMSGRLLGAFRVPVALMRHLGNELGIAAPEVASRAMTERGRTLF  
DHOQVACTVLGFQWMSHQRSLVRELDEVARCADRDQLLVRARQWLYKNKLVIVHE  
RAIRTLIAAALQLEVEGTGTAAASVDGTPATLDRWRASVSELRPDGQTQQSWLWAAPAK  
HSTROISEVLERIDLLYTLDVHKHLADIPLLIRRYARLVSRPSPAGAKIKEPARTV  
EVACFLRYCLFTTTDOLIMVQRRIADLRQAADVPATVNNMAAMYKTLIGELVALSA  
QGAVPDALRLARLEALITETQKRPSPRSASLIREGLIDGIRPVRSLLVAIAKLQWAT  
GEPATLEYLAKLQALYDKSRKLPVEVVAFSLGNWQVSISSPDRERAFQALEVATLF  
ALRRVANGSVWIEHSLSPFRARLFTDERWQAESKKHYARUSLSPKAATFTKLPLLA  
RVTAGVDVAAAARSGVLRVDDHLHSLPAEDDEDPVTKLRAALDRHRIIGEVQLPEVI  
LAVDAQVRFWSIMLGRPRSTDELLMYYAGTMAHGTSLTAVECARMIPQASATLIRQA  
MRWARDERLSQACQAVLEFMORHPAAATWGRSDLASDDMMSETTKRWQARLDPRR  
NTESIGIYSHVKORWGLFHAQPFVLRERQAGVIEGVIROEKLSTSLAVDTHGYTDF  
AMSHARILGFDLPRCLKELKQRLVFRPGTKVPAEIAAVCEANVDVALIEKHWSLVH  
LAASVMSGHASAVAALAFGSAQCDPIYEAGVQLGRLLRTAFLLAFVDFKAFNELR  
RVLNRGPAVNALKRAIYTGRISPAQAQRVDEMQAADALSLMANIVMAWNTSQMAVL  
DRWSNRQVTPPELIGKIAPTRLBSINLRGVFRFPVDRYADQILPSRPNASITGTNG"  
4564. 5238  
/gene="dehH2"  
4564. 5238  
/gene="dehH2"  
/function="dehalogenation"  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="haloacetate dehalogenase H-2"  
/protein\_id="BAC81974.1"  
/db\_xref="GI:34013301"  
/translation="MKKIEAIPDMYGTLYDVHVSVDACEKQYPGKGDISVLWRQKQ  
LEYNLACLNGQVIKFEATANALTYTCNMKLCDDGSSAMRLTEEYLRKLPFPEVRG  
ALRALRQGRMLAILNSGTSTETHDVVHNSGVEGEPEHLISVDSARAYKPHPLAYELG  
EEAFGISRESILFVSSNPWDVSGAKAFGYQCVINRYGFAFDELQGTPTFVPMDAI  
VHLIAY"  
/complement(5446. .6012)  
/gene="orf1"  
/complement(5446. .6012)  
/gene="orf1"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC81975.1"  
/db\_xref="GI:34013302"  
/translation="MSEPEKIVHEEKAQDFQSQSMYGDYLLHDLMLKAQHPLSPAHD  
EMLFIVQHTSELWMKMLHMLHMAATRIATEQLPDAFKMLARVSRINEQLVGARVL  
STMTPEYSAMRPLASSGGSQAQYICIEFALGNKNAAMLKPHARADLLALVDAAI  
RRTLSLYDEALRLARRGLAVPPGYVERN"  
6681. .9884  
/note="IS1071"  
6826. .9741  
/gene="tnpA"  
6826. .9741  
/gene="tnpA"  
/function="transposition"  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="transposase"  
/protein\_id="BAC81976.1"  
/db\_xref="GI:34013303"  
/translation="MQGWHHTTFLGMRGLPROISDFENKAFTEFGAERDAINARRGDS  
HKLGLALHIGFLRMSGRLLGAFRVPVALMRHLGNELGIAAPEVASRAMTERGRTLF  
DHOQVACTVLGFQWMSHQRSLVRELDEVARCADRDQLLVRARQWLYKNKLVIVHE  
RAIRTLIAAALQLEVEGTGTAAASVDGTPATLDRWRASVSELRPDGQTQQSWLWAAPAK

HSTRQISEVLRLIDLLYTLVDVHKHLADIPIDLILRRYARRLVRSPPSAGAKIKEPARTV  
 EVACFRYCLFTTDDILMVQRRIADLRQAADVPATVNWAAKYTLIGELVALSA  
 QGAPDAELRRLALITETQKPPSRASLAREGLDGIIRPVRSLLVAIAKLPAQAT  
 GEPALLEYLKLQALYIKSRKLPVENVAPSLGNWQVSISSPDRERAFQALEVATLF  
 ALRRVNGSVWTEHLSFRGRARLFTDERWOAESKHYARLSLPSKAATFLKPLLA  
 RVTAGDVAAAASGVLRVDDDELHLSPLPAEDDEPVTKLRAALDHRIGEVOLPEVI  
 LAVDAQVRSWIMLGRPSTDBLLMVYAGINVAHGTSLTAVECARMIPOLSATSIRQA  
 MRWARBERLSQAQVAFEMORHPITAAWGRSDLASDMMSEMTTKRWYQARLDPRR  
 NTPSIGIYSHVKDWGIEHAQPFVLERQAGVAIEGVIRQEKLETSQALADVTHGYTDF  
 AKSHARLLGDLPRKELKQRLHFPVPRGTQVPAETAANVCEANVDVALLKEKWDLSLVH  
 LAASVGHASAAVALARFGSAAGQDPIYEAGVOLGRLLKRTAFADYFVKDAFRNELR  
 RLNRGEAVNALKRATIGRISPAQAKRVDEMOAVADLSLMANIYMWANTSQOMQAVL  
 DVMNRQVPIPPLEGIKPIATRLIESINLRGVFRFPVDYADQILPSPRNASITGTNG"  
 10043..10801  
 /gene="orf2"  
 10043..10801  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC81977.1"  
 /db\_xref="GI:34013304"  
 /translation="VVDTHFKOHFSSGAGYAARYPTYTELVDALANISPLGERALD  
 CGCGTQSLVLLERAEVAVDASAAQIDKQAPYDGVYTRALAAESGLPDASVDLI  
 TVAAAHWLDLDFYAEVQRVARSQAVALITYVLHVEGAVDGTQRFYDTIGEPY  
 PTERREVEDGYRSLDPFPFEVALPPLAIEVOWRLDLDLIGYLNWTSVAKAEKALGNSP  
 VDAAETLHKWGDPALHRVTWPLSVRAGHVQK"  
 10798..11235  
 /gene="orf3"  
 10798..11235  
 /gene="orf3"  
 /function="unknown"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC81978.1"  
 /db\_xref="GI:34013305"  
 /translation="MKDAWAKNHPILGLRHSQTLRVDDSLTVPAVSAFTGFSMDPP  
 VFATYVNGFVEWACIEALRPLYPASQRTVGHVNLHSAATPVGMQVTAELIEVE  
 GKRLTFKVLCDREVDCIEGRHERFIVEAQFIRRVTSKGRG"  
 11276..12160  
 /gene="deh1"  
 11276..12160  
 /gene="deh1"  
 /function="dehalogenation"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="haloacetate dehalogenase H-1"  
 /protein\_id="BAC81379.1"  
 /db\_xref="GI:34013306"  
 /translation="MDFPFGKNSVTVTVDGVDIAYTVSGEGPVLMLHGPPQNRAMWAR  
 VAPQAEHTTVVCAIDRGYDSDPKLPDRSNYSRPTFAHQDLQVYMRHGFERFHLV  
 GHDRCGRTHRMALDHPREALVSLTVMIDIVTPYMPMNTNRLVAASVHWYFQPEPF  
 PEHMGDQDPFFYETCLFCWGATKVSDFDQMLNAVRESWRNPAMTHGSCSYRAAAT  
 IDLEHDSAIQRKVECPITLVFYGSKQMGQLFDIPAEWAKRCNNTYNASLPGHFVD  
 QPFAETSEILKFLARNG"  
 complement(12335..13306)  
 /gene="orf4"  
 complement(12335..13306)  
 /gene="orf4"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC81980.1"  
 /db\_xref="GI:34013307"  
 /translation="MLIALHKQARTTPAVRAEIAASDEPASILAQRFGITEQTVYKWK  
 KRDFADRGHTARLQTLVPAQETVVVHLRLTLPLDLDLAVTRREFICPVRSGL  
 DCLRLRHGAGNALKAPQPAVAHAKFKSEYGVHMDVKYLPMQDESSRYLYLVAI  
 DRATRVYQLKSNKTAASQAQFLKALHKACPIRINKLITDNGKFTDLRLFASKREP  
 SGNHFDQDLGIEHRLUTKPTKPTNGWVERFNGRIADVLKTRHFNREDMEETLL  
 RYVALYNHQLPOSALKSSTPMQAMQDWYQTHPHLPHKRPYDRTGDS"  
 complement(13685..14581)  
 /gene="orf5"  
 complement(13685..14581)  
 /gene="orf5"

Query Match 14.4%; Score 46.2; DB 1; Length 67066;  
 Best Local Similarity 47.7%; Pred.No.33;  
 Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;  
 QY 6 CCACAAGCGCACTAAACGCGACGCCATCCGCGAGCGGTCAACGCCCCCGGTGCGAA 65  
 Db 60318 CGACAAGGCCAGGAGAGGGCCACGCCAGCCAGCGTGGAGGCGCTCGGGCCGA 60259  
 QY 66 TCGCGTTGGCGGCAACATGGTTGGCGGCCAGCGTTCGTCGCGGAGCAGGCGCGC 125  
 Db 60258 CTTCCGCGCGCGGCAACAGCGGTACCGCGAGATCGAGCGCGCGCGGCGAGCTGCGTGC 60199  
 QY 126 CCAACGCGGCGCGACCTCGAGGCTATCCGCGAGCGTACGCGCGAGATGTTGCGCACATC 185  
 Db 60198 CGACCTGGAGCGGGGCAACAGCGACCGAGCGCGCGCGCTCAGCGCGAGCAGCG 60139  
 QY 186 ACACGAATTCGACGACGACACACAGCCGAACTGGCGTGTTCATGTCATCTCGACGATGA 245  
 Db 60138 GAAGGCCACCGAGGCGCGCGCGCGAGCGCGGACAGGTGCGCGGAGCTGCTGAAGGT 60079  
 QY 246 GCAGCGCGCGCTTGGCGGGGATGAAGCTCGCTGGCATCCG 288  
 Db 60078 GCAGGCCAAGCGCGGCGCGCGAGCGCGAGCGCCACCGAGGAGCAG 60036  
 RESULT 9  
 AY365053/c 87688 bp DNA circular BCT 04-JUN-2004  
 LOCUS Wautersia eutropha strain JMP134 plasmid pJP4, complete sequence.  
 DEFINITION Wautersia eutropha strain JMP134 plasmid pJP4, complete sequence.  
 ACCESSION AY365053  
 VERSION AY365053.1 GI:39777443  
 SOURCE Wautersia eutropha  
 ORGANISM Wautersia eutropha  
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 Burkholderiaceae; Wautersia.  
 Trefault,N., De la Iglesia,R., Molina,A.M., Manzano,M., Ledger,T.,  
 Perez-Pantoja,D., Sanchez,M.A., Stuardo,M. and Gonzalez,B.  
 Genetic organization of the catabolic plasmid pJP4 from Ralstonia  
 eutropha JMP134 (pJP4) reveals mechanisms of adaptation to  
 chloroaromatic pollutants and evolution of specialized  
 chloroaromatic degradation pathways  
 Environ. Microbiol. 6 (7), 655-668 (2004)  
 REFERENCE 2 (bases 1 to 87688)  
 Trefault,N. and Gonzalez,B.  
 Direct Submission  
 Submitted (11-AUG-2003) Departamento de Genetica Molecular Y  
 Microbiologia, Facultad de Ciencias Biologicas, Pontificia  
 Universidad Catolica de Chile, Casilla 114-D, Santiago, Chile  
 Location/Qualifiers  
 1..87688  
 /organism="Wautersia eutropha"  
 /mol\_type="genomic DNA"  
 /strain="JMP134"  
 /db\_xref="taxon:510"  
 /plasmid="pJP4"  
 join(55727..87688,1..8678)  
 /note="incP-lb beta backbone"  
 complement(165..818)  
 /gene="trn"  
 complement(165..818)  
 /gene="trn"  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="trn"  
 /protein\_id="AA31019.1"  
 /db\_xref="GI:39777444"  
 /translation="MKRTIPGSGTESRATKPARVSAPATLGGAAFGASREDTGADLL  
 EAAQAAETEQATLEAAPEVQSYPETLALYVQAKHDOVEHIEDLENLIDROQARLQ  
 TQASAPGRSLSPGSKRANQQAQQAARLQTLHARLEAVREIKEGMGLHSPKIEELAT





/gene="k1cB"  
/note="ORF11"  
/codon\_start=1  
/transl\_table=11  
/product="stable inheritance protein"  
/protein\_id="AAK31030.1"  
/db\_xref="GI:39777455"  
DAALMRGDGAAEAASDRYEAVIWKNGGTHVSGSMADHAPGQVIERHCRAVPGDVL

Query Match 14.4%; Score 46.2; DB 1; Length 87688;  
Best Local Similarity 47.7%; Pred. No. 31;  
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 6 CCACAAGCGCACTAAACGCGCAGCAGCATCGCGCAGGCTCAACGCCGCCGCGTCCGAA 65  
DB 1848 CGACAGGCCACGAGAGGCCACGCCAGCGCCGACGAGTGGAGGCCCTCGGCCCGA 1789  
QY 66 TCCTGTTGGCGCGCAACATGTTGGCGCGCGACGTTTCGTCGCCGCGAGCAGCGCCGCG 125  
DB 1788 CCTCGCGCGCGCAACAGCGCTACCGCGGAGATCGAGCGCGCGCGGAGCTGCGTGC 1729  
QY 126 CCAACGCGCAGCGCACTCGAGGCTATCCCGAGCGTACCGCGAGTGGTGGCGACATC 185  
DB 1728 CGACCTGGAGCGCGCGCAACAGCGACCGACCGAGCGCGCGCGCTCAGCGCCGAGCAG 1669  
QY 186 ACACGAAATCGACGACGACAGCGAACTGGCGCTGTTGTCGATGCATCTCGACGATGA 245  
DB 1668 GAAGCCACGAGCGCGCGCGCGCGGACGCGGACGAGTGGCGCGGAGCTGTGAAGT 1609  
QY 246 GCAGCGCGCGCTTGAGCGCGGATGAAGCTCGCTCGCATCCG 288  
DB 1608 GCAGCGCAAGCGCGCGCGCGCGAGCAGCGCCACCGAGGACG 1566

RESULT 10  
U66917  
LOCUS 108845 bp DNA circular BCT 03-JUL-2003  
DEFINITION Pseudomonas sp. ADP atrazine catabolic plasmid PADP-1, complete  
ACCESSION U66917  
VERSION AF017572  
KEYWORDS U66917.2 GI:13937422  
SOURCE Pseudomonas sp. ADP  
ORGANISM Bacteria; Proteobacteria.  
REFERENCE 1 (bases 34964 to 36388)  
de Souza,M.L., Wackett,L.P., Boundy-Mills,K.L., Mandelbaum,R.T. and Sadowsky,M.J.  
Cloning, characterization, and expression of a gene region from Pseudomonas sp. strain ADP involved in the dechlorination of atrazine  
J. Appl. Environ. Microbiol. 61 (9), 3373-3378 (1995)  
MEDLINE 96035669  
PUBMED 754846  
REFERENCE 2 (bases 43552 to 46212)  
Boundy-Mills,K.L., de Souza,M.L., Mandelbaum,R.T., Wackett,L.P. and Sadowsky,M.J.  
The atzB gene of Pseudomonas sp. strain ADP encodes the second enzyme of a novel atrazine degradation pathway  
Appl. Environ. Microbiol. 63 (3), 916-923 (1997)  
MEDLINE 97208208  
PUBMED 9055410  
REFERENCE 3 (bases 69775 to 71867)  
Sadowsky,M.J., Tong,Z., de Souza,M. and Wackett,L.P.  
AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes  
J. Bacteriol. 180 (1), 152-158 (1998)  
MEDLINE 98083068  
PUBMED 9422605  
REFERENCE 4 (bases 43552 to 46212)  
Boundy-Mills,K.L., de Souza,M.L., Mandelbaum,R.T., Wackett,L.P. and Sadowsky,M.J.

TITLE Direct Submission  
JOURNAL Submitted (15-AUG-1996) Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108, USA  
REFERENCE 5 (bases 69775 to 71867)  
AUTHORS Sadowsky,M.J., Tong,Z., de Souza,M.L. and Wackett,L.P.  
TITLE Direct Submission  
JOURNAL Submitted (05-AUG-1997) Soil, Water, and Climate Dept., University of Minnesota, 1991 Upper Buford Circle, 439 Borh, St. Paul, MN 55108, USA  
REFERENCE 6 (bases 1 to 108845)  
AUTHORS Martinez,B.M., Tomkins,J., Wackett,L.P., Wing,R. and Sadowsky,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-JAN-2001) Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Ave., 140 Gortner Labs, Saint Paul, MN 55108, USA  
REMARK Nucleotide sequence updated by submitter  
COMMENT On or before May 3, 2001 this sequence version replaced gi:2736100, gi:1732205.  
FEATURES  
Source Location/Qualifiers  
1..108845  
/organism="Pseudomonas sp. ADP"  
/mol\_type="genomic DNA"  
/strain="ADP"  
/db\_xref="taxon:47660"  
/plasmid="PADP-1"  
/note="atrazine catabolic plasmid"  
complement(869..1183)  
/gene="orf1"  
complement(869..1183)  
/gene="orf1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK50234.1"  
/db\_xref="GI:13937423"  
/translation="MSYAVVFSPEAEQALRYIAAASPDVAARYTEAIVTYCES LRTFTPLRGTRMDVRPGLRITNYRKRVIADFVAGDLVSIIGVFYGGQDYESILHDDA DD"  
complement(1180..1461)  
/gene="orf2"  
complement(1180..1461)  
/gene="orf2"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK50235.1"  
/db\_xref="GI:13937424"  
/translation="MEARMTTQQLSITLPNDMADVTKTKVKGYSASESEVIRDLR ALLARDRAVESWLNQVGPAYDALKADPSRAVTADQVRARLAAEHAQAQ"  
1618..2046  
/gene="k1cA"  
1618..2046  
/gene="k1cA"  
/note="protein involved in plasmid stable inheritance"  
/codon\_start=1  
/transl\_table=11  
/product="k1cA protein"  
/protein\_id="AAK50236.1"  
/db\_xref="GI:13937425"  
/translation="MNTQDQPVTSALVAEQRLDFPTFGPRLLMRGALVYAMLR LCIERYGAYWHYITLSDGFLAPDLAERLIEVDGNGFRGELSADAAGIVATLFAIG QLAARIAGTDAADALIDRYHFLRGPAAGHPAAAIYRAID"  
2098..2517  
/gene="k1cB"  
2098..2517  
/gene="k1cB"  
/note="protein involved in plasmid stable inheritance"  
/codon\_start=1  
/transl\_table=11  
/product="k1cB protein"  
/protein\_id="AAK50237.1"  
/db\_xref="GI:13937426"

```
gene /translation="MAKQTKQAPAGQDWSAPAAELIAELPADRDLGLAALAAVAWVEI
CDS DAAYVRGDGAALAGDRYEAI IWKLGNTNFGWADDEAAGRVIERHCAAVPGDVEL
3427. 3684 /db_xref="GI:13937432"
/gene="korC" /product="KorC protein"
3427. 3684 /db_xref="GI:13937428"
/gene="korC" /note="transcriptional repressor"
/codon_start=1
/transl_table=11
/product="KorC protein"
/protein_id="AAKS0238.1"
/db_xref="GI:13937427"
/translation="MTNDANIRLECLPAERWAQPSGSEVREVLRLAGLTGGKAQVL
GLPGKQDRTIRWVGEDTPIPYAAWALLCDVAGLGLIWEA"
3844. 4080 /db_xref="GI:13937433"
/gene="klea" /product="KleA protein"
3844. 4080 /db_xref="GI:13937428"
/gene="klea" /note="protein involved in plasmid stable inheritance"
/codon_start=1
/transl_table=11
/product="KleA protein"
/protein_id="AAKS0239.1"
/db_xref="GI:13937428"
/translation="MSKNKIMPWVDALPNVEATDFQARRDQIEATMAEAAELVQABE
LRGKAYFAALSLEASAKGESSQAVEQAKRSVGN"
4139. 4354 /db_xref="GI:13937429"
/gene="kleB" /product="KleB protein"
4139. 4354 /db_xref="GI:13937428"
/gene="kleB" /note="protein involved in plasmid stable inheritance"
/codon_start=1
/transl_table=11
/product="KleB protein"
/protein_id="AAKS0240.1"
/db_xref="GI:13937429"
/translation="MSKGKIEIETCCRRCGSIRTLSHTIIGADAREKFGSICGGC
ITPDEDNELTEMLAAAVRRMSGATLQ"
4442. 4828 /db_xref="GI:13937430"
/gene="kleG" /product="KleG protein"
4442. 4828 /db_xref="GI:13937430"
/gene="kleG" /note="protein involved on plasmid stable inheritance"
/codon_start=1
/transl_table=11
/product="KleG protein"
/protein_id="AAKS0241.1"
/db_xref="GI:13937430"
/translation="MLPSGSIQHSRNRVEQGMNSIVKPPRASKPPAPEVQPAAPA
AAPAKPAEGRLVAGLVKFWVATLVWPKLVLAITFFQFVRLYHNTPGVVA
GWSFLAYPAALTAITYFVSIYKPKGL"
4830. 5360 /db_xref="GI:13937431"
/gene="kleF" /product="KleF protein"
4830. 5360 /db_xref="GI:13937431"
/gene="kleF" /note="protein involved in plasmid stable inheritance"
/codon_start=1
/transl_table=11
/product="KleF protein"
/protein_id="AAKS0242.1"
/db_xref="GI:13937431"
/translation="MAKOTLPYPPGFBPTTGRVAVLVREYADSLNGDAPYWYSAQ
SEELQDPRWLVGVDPHVGGSDVQVCFASGTRTVGLMTFFLSAAHAQOLIDAKGE
BLAQRLATLAVIADGLGHPAKALRIEAKVEGRPAVFYDQDQATLCACAVSDHWQRAR
ATAATASAIQKATNF"
5104. 5367 /db_xref="GI:13937432"
/gene="kleG" /product="KleG protein"
5104. 5367 /db_xref="GI:13937432"
/gene="kleG" /note="protein involved in plasmid stable inheritance"
/codon_start=1
/transl_table=11
```

```
gene /product="KleG protein"
CDS /protein_id="AAKS0243.1"
/db_xref="GI:13937432"
/transl_table=11
/translation="MRHSSLTTPRGKSWPCAPPWPSRTGWACLPRRCASRRWRPAGR
PCSTTRTGPRAPARWIPYTGDRGRPPQHRPSTRREQIFSR"
5461. 6537 /db_xref="GI:13937433"
/gene="incC" /product="InclC protein"
5461. 6537 /db_xref="GI:13937433"
/gene="incC" /note="inclusion membrane protein"
/codon_start=1
/transl_table=11
/product="InclC protein"
/protein_id="AAKS0244.1"
/db_xref="GI:13937433"
/translation="MGAIHEETANRSPIPDHOGAGDRAADHRHSARRAGRWAPPGV
CFPAGADQGGVAGGQSRMSSRTAPRGLRASDGTAGASSVHRQVSGSRORBTGT
QVMKLTVAIQKGGQKTFATFCHLAFLEGLRVLRAVIDLDTQGNASFTLSAYQSGYL
```

```
Query Match 14.4%; Score 46.2; DB 1; Length 108845;
Best Local Similarity 47.7%; Pred. No. 30;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
Qy 6 CCACAAGCGCACTAAACCCAGCCAGCCATCGCCGCGAGGGCTCAACGCCCGCGTGGAA 65
Db 8311 CGACAAGCGCCAGAGAGGGCCACGCGCCAGCGCCAGCGTGGAGGCCCTCGCGGCCGA 8370
Qy 66 TCGGTTTGGCGGCAACATGTTGGCGCGGCGAGCTTCCTCCGCCGAGAGCGCGCGC 125
Db 8371 CCTCGCGCGCGAACCAGCCGATCCGCGCGAGATCGAGCGCGCGCGCGAGTGTGCTGC 8430
Qy 126 CCAACGCGCAGCGGACCTCGAGGCTATCCGCGAGCGGTACCGCGAGATGTTGGCGACATC 185
Db 8431 CGACCTGGAGCGGCGAACCAGCGCCAGCGCGCGCGCGCTCAGCGCGAGCAGCG 8490
Qy 186 ACAGAAATCGACAGCACAGCGGAACTGGCGCTGTTGCGATGATCTCGAGATGA 245
Db 8491 GAAGGCCACCGAGCGCGCGCGCGCGAGCGGCGGACCGTGGCGCGGAGCTGGTGAAGT 8550
Qy 246 GCAGCGCGCGTTCGAGCGGGATGAAGCTCGGCTGGCATCCG 288
Db 8551 GCAGCCAAAGCGCGCGCGCGCGAGCGCGCCACAGAGCAG 8593
```

```
RESULT 11
AE016924/c 301457 bp DNA linear BCT 14-OCT-2003
LOCUS Chromobacterium violaceum ATCC 12472 section 15 of 16 of the
DEFINITION complete genome.
ACCESSION AE016924 AE016825
VERSION AE016924.1 GI:34105231
KEYWORDS
SOURCE Chromobacterium violaceum ATCC 12472
ORGANISM Chromobacterium violaceum ATCC 12472
REFERENCE 1 Chromobacterium violaceum ATCC 12472
AUTHORS Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
CONSRM Neisseriaceae; Chromobacterium.
TITLE 1 (bases 1 to 301457)
JOURNAL Brazilian National Genome Project Consortium
MEDLINE The complete genome sequence of Chromobacterium violaceum reveals
PUBMED remarkable and exploitable bacterial adaptability
AUTHORS Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)
2282880
PUBMED 14500782
2 (bases 1 to 301457)
AUTHORS Vasconcelos,A.T.R., de Almeida,D.F., Almeida,F.C., de
TITLE Almeida,L.G.P., de Almeida,R., Goncalves,J.A.A., Andrade,E.M.,
JOURNAL Antonio,R.V., Araripe,J., de Araujo,M.F.F., Filho,S.A., Azevedo,V.,
MEDLINE Batista,A.J., Bataus,L.A.M., Batista,J.S., Belo,A., vander Berg,C.,
PUBMED Blamey,J., Bogo,M., Bonato,S., Bordignon,J., Brito,C.A.,
AUTHORS Brocchi,M., Burity,H.A., Camargo,A.A., Cardoso,D.D.P.,
REFERENCE Carneiro,N.P., Carraro,D.M., Carvalho,C.M.B., Cascardo,J.C.M.,
Cavada,B.S., Chueire,L.M.O., Pasa,T.B.C., Duran,N., Fagundes,N.,
```

Falcao,C.L., Pantinatti,F., Farias,I.P., Felipe,M.S.S.,  
 Ferrari,L.P., Ferro,J.A., Ferro,M.I.T., Franco,G.R.,  
 Freitas,N.S.A., Furlan,L.R., Gazzinelli,R.T., Gomes,E.A.,  
 Goncalves,P.R., Grangeiro,T.B., Grattapaglia,D., Grisard,E.C.,  
 Guimaraes,C.I., Hanna,E.S., Hungria,M., Jardim,S.N., Laurino,J.,  
 Leao,L.C.T., Passarella,L., Lima,A., Loureiro,M.F., Lyra,M.C.P.,  
 Macedo,M., Madeira,H.M.F., Manfio,G.P., Maranhao,A.Q.,  
 Martins,V.M., Mauro,S.M.Z., de Medeiros,S.R.B., Meissner,R.V.,  
 Menck,C.F.M., Moreira,M.A.M., Nascimento,F.P., Nicolas,M.F.,  
 Oliveira,J.G., Oliveira,S.C., Paixao,R.F.C., Parente,J.A.,  
 Pedrosa,F.O., Pena,S.J.D., Perreira,J.O., Perreira,M.,  
 Pinto,L.S.R.C., Pinto,L.S., Porto,J.I.R., Potrich,D.P.,  
 Neto,C.E.R., Reis,A.M.M., Rigo,L.U., Rondinelli,E., dos  
 Santos,E.B.P., Santos,F.R., Schneider,M.P.C., Senanez,H.N.,  
 Silva,A.M.R., da Silva,A.L.C., Silva,D.W., Silva,R., Simoes,I.C.,  
 Simon,D., Soares,C.M.A., Soares,R.B.A., Souza,E.M., Souza,K.R.L.,  
 Souza,R.C., Steffens,M.B.R., Steindel,M., Teixeira,S.R.,  
 Urmenyi,T., Vettore,A., Wassem,R., Zaha,A. and Simpson,A.J.G.  
 Direct Submission  
 Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de  
 Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ  
 25651070, Brazil  
 FEATURES  
 source  
 1. 301457  
 /organism="Chromobacterium violaceum ATCC 12472"  
 /mol\_type="Genomic DNA"  
 /strain="ATCC 12472"  
 /db\_xref="taxon:243365"  
 121. 1386  
 /gene="goag"  
 /locus tag="CV3926"  
 121. 1386  
 /gene="goag"  
 /locus tag="CV3926"  
 /EC number="2.6.1.19"  
 /notes="identified by sequence similarity; putative; ORF  
 located using Glimmer/Genemark/Blastx/COG160"  
 /codon\_start=1  
 /transl\_table=11  
 /product="4-aminobutyrate transaminase"  
 /protein\_id="AAQ61587.1"  
 /db\_xref="GI:34105232"  
 /translations="MSQQQLMRKSDATPRGVGVCFFADRAKNAEVDWTDGRRYID  
 FAGIGVNLTHLPKVAEVAEOLRFSGTQVVPVYESYISLAEKNALAPGKHAK  
 KTAFFTEAEVNAVKIARAATGRQVIAFGAFHGRITLMGALTGVKPYKLGFGP  
 FPSDVHAPPTLHGVSQDSLVGLEKLPKFDIDPARVAAIIVPEVQEGGFYAAPA  
 EWKALRELCDHGLIIDADEVOSGFARTGKLLASHFPAVDPLITWAKSLAGGFPLS  
 AVTGRAEVDMPAPGGLGGTYAGNPLAAAHAVLDVIDSEKLAERAQRLGALKKELI  
 NGLKSAVQIAEVRGLGAMTAVFNQPSHQPPDFPAKKVQOQHALANGLLILTCGIYA  
 NVRFPLPTIEDIEFAELGKLEAALKA"  
 1456. 2916  
 /gene="gabd"  
 /locus tag="CV3927"  
 1456. 2916  
 /gene="gabd"  
 /locus tag="CV3927"  
 /EC number="1.2.1.16"  
 /notes="identified by sequence similarity; putative; ORF  
 located using Glimmer/Genemark/Blastx/COG1012"  
 /codon\_start=1  
 /transl\_table=11  
 /product="succinate-semialdehyde dehydrogenase [NAD(P)]"  
 /protein\_id="AAQ61588.1"  
 /db\_xref="GI:34105233"  
 /translations="WNLQDPSLLRQCYLGGEWLDADGGHIAVHNPAATGEKIGQVP  
 KGRABEAEVAAARAPPAWAKSAKERAAILRFWEDLIIAHOQDLAILTSEQGRF  
 LAEAKGETAYGASTVETAEAKVYGDIVPGPADRRVLVTEPIGVTAIINWPP  
 SAMITKRAAPALAGCTVVRPASQTPLSALAEALRAGIPAGVSVITGASSELG  
 ALVTGSDTVKFTSGTEVGRKLIQAQCACTGVKVSMELGNAFFIVDDADDAVEE  
 GALLISKYNAQGTVCNARIYVQGVYDAFAEKFAAAVATLKVNGLEAGVSGQLID  
 HNAVKVEITADALAKGRRIAGKRGHALGTFTEPTLIADVTRDMKVAAREETFGPL  
 APLFPKTEFEERAILANDTEFLASIFYVARDVGRIFRVAEGLEYGMVAVNSGLLSNEA  
 APFGVKOSGLGRESKYGIEDYLEIKVLLGLDR"

3010. 3330  
 /locus tag="CV3928"  
 3010. 3330  
 /locus tag="CV3928"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAQ61589.1"  
 /db\_xref="GI:34105234"  
 /translations="MKYNHIGTSGRFDNEIDLPHLKMVSDHQDNAFGIOWORYWQ  
 DAPYELVKKRPHIAFEVEDLAQLEGHKLLIAPNSPGLITVAFIEVNGAPVELMQY  
 DKNN"  
 complement (3435. 3674)  
 /gene="rRNA5S"  
 /locus tag="CVrRNA-5s-7"  
 complement (3435. 3674)  
 /gene="rRNA5S"  
 /locus tag="CVrRNA-5s-7"  
 /product="rRNA-5s"  
 /notes="identified by sequence similarity; putative"  
 complement (3675. 6614)  
 /gene="rRNA23S"  
 /locus tag="CVrRNA-23s-7"  
 complement (3675. 6614)  
 /gene="rRNA23S"  
 /locus tag="CVrRNA-23s-7"  
 /product="rRNA-23s"  
 /notes="identified by sequence similarity; putative"  
 complement (6813. 6888)  
 /locus tag="CVtRNA-Ala-TGC-7"  
 complement (6813. 6888)  
 /locus tag="CVtRNA-Ala-TGC-7"  
 /product="tRNA-Ala"  
 complement (6975. 7051)  
 /locus tag="CVtRNA-ile-GAT-7"  
 complement (6975. 7051)  
 /locus tag="CVtRNA-ile-GAT-7"  
 /product="tRNA-ile"  
 complement (7295. 8768)  
 /gene="rRNA16S"  
 /locus tag="CVrRNA-16s-7"  
 complement (7295. 8768)  
 /gene="rRNA16S"  
 /locus tag="CVrRNA-16s-7"  
 /product="rRNA-16s"  
 /notes="identified by sequence similarity; putative"  
 8828. 9085  
 /locus tag="CV3929"  
 8828. 9085  
 /locus tag="CV3929"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAQ61591.1"  
 /db\_xref="GI:34105236"  
 /translations="MCKYSAPAKHLHLSVHVKQCRSLRRRCQTVSFPLFRSLRQ  
 LARRTIRDPKPVNTFREKNLRKMPILFQNNNKLKRXHLC"  
 9220. 10809  
 /gene="nadB"  
 /locus tag="CV3930"  
 9220. 10809  
 /gene="nadB"  
 /locus tag="CV3930"  
 /EC number="1.4.3.16"  
 /notes="identified by sequence similarity; putative; ORF  
 located using Glimmer/Genemark/Blastx/COG0029"  
 /codon\_start=1  
 /transl\_table=11  
 /product="L-aspartate oxidase"  
 /protein\_id="AAQ61592.1"  
 /db\_xref="GI:34105237"  
 /translations="MLKFDVLLIGSLAGMTLALHAETRKVGLITKRDLLGNSYTA  
 QGGIAAVLTDSDVEDHIDHTLVAGLGCNEKATRFIVHSKEAIDWLKLGVPFTRD"

DEHOTGHLTREGGSHRIIHAADATGAATVSTLQKIKAPHNITELHYIAIDLIT  
SOKLGKRAYGVVLDKNGDEVVTTAAOHTILASGAGKVLYTTPDTPATGDIAT  
MGWAGCVGNMERTQHEPTCTLYHPAKSFLISEAVRGEGLKLPDGTREMPGDPR  
GELAPDVVARAIDPEMKHGLDVCYLDISYKPAQFIKEHPNFIYACLEGLDITQO  
PIPVPAVHFTCGGIVTLAGRTVDGDIYAVEVACTGLHGANRLANSLECLVIGK  
AAADILDAADSIKLPATPAWDSRVTOPDEEVIWISHNWDLRAMVDYGVIRTKRL  
EALRHLIDLLKEIDEYVSHFHVNDLIELNLVQTAELIVRSALLRKESRGLHSRD  
YPNLLPEAAETVLAADPLGG"  
11124. 12206  
/locus\_tag="CV3931"  
11124. 12206  
/locus\_tag="CV3931"  
/EC number="3.2.1.132"  
/note="identified by sequence similarity; putative; ORF  
located using GeneMark/Blastx"  
/codon\_start=1  
/transl\_table=11  
/product="probable chitosanase A"  
/protein\_id="AAQ61593.1"  
/db\_xref="GI:34105238"  
/translation="MKYAAKRLSPSATLGLALLAAGSAAAGTAGSPEROTVRL  
PDCAANSASPYPQAGSVASRGVNYTAAFTQGNPPEOGAOWAGKRCRPAQAKAA  
DHDANFPATLTKFLKANTGLDGEQNDIMKLVNPEQDSLEWTFYGCENIGKRGY  
TIGIFGATVGGNDEGPDPTLFKEFDASSGASNPISAGGLARAGVHSGMGKILKIS  
DSAKVCDKIGSLQNPAMDAMNTFYKYIYQSVQOARQGFSSALTIGSFVDTAL  
NQGATGSGTLQGLLSRSGSGSDEKTFMTAFYQSRKIVDTNDYNQPNPKRVKQWS  
TLNNGETDLKNAADAAVQKVTWEMK"

Query Match 14.1%; Score 45.4; DB 1; Length 301457;  
Best Local Similarity 51.8%; Pred. No. 35;  
Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 104 CGTCCGCGAGCAGCGCGCCCAACGCGCGGACCTCGAGGCTATCCGCGAGCGT 163  
DB 170537 CTTGGCGGAAACCGGCCAACGAGGCGCGCTGCGCGCGCTGCGCCACCGAGCTGGCGG 170478

QY 164 AGCGCGAGATGTGCGGCACATCACAGAAATCGACGACGACACACGACCGAATCGGCGCTGT 223  
DB 170477 GCCTCGACGCGCGGGATCCAGGCGGAACCTGGAAGACGAAACGCGCTGGGAGCGCCAGC 170418

QY 224 TGTGATGATCTGACAGATGAGCGCGCGCTTGAGCGGGGAGTGAAGCTCGGCTGCG 283  
DB 170417 TGTGGAAGATGCCAGACACACCGAGCGCTGCGCGGAGCATGGCTTCATGCGCGCGCTGCG 170358

QY 284 ATCCGTATCACTTCCCGCA 302  
DB 170357 GCATATTTCCCGCGCA 170339

RESULT 12  
AP005033 320150 bp DNA linear BCT 10-MAY-2003  
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 13/30.  
DEFINITION AP005033 BA000030  
VERSION AP005033.1 GI:29606546  
KEYWORDS Streptomyces avermitilis MA-4680  
SOURCE Streptomyces avermitilis MA-4680  
ORGANISM Streptomyces avermitilis MA-4680  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Omura.S., Ikeda.H., Ishikawa.J., Hanamoto.A., Takahashi.C., Shinose.M., Takahashi.Y., Horikawa.H., Nakazawa.H., Osonoe.T., Kikuchi.H., Shiba.T., Sakaki.Y. and Hattori.M.  
Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
MEDLINE 11572948  
PUBMED  
REFERENCE 2  
AUTHORS Ikeda.H., Ishikawa.J., Hanamoto.A., Shinose.M., Kikuchi.H.,

Shiba.T., Sakaki.Y., Hattori.M. and Omura.S.  
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306  
MEDLINE 12692562  
PUBMED  
REFERENCE 3 (bases 1 to 320150)  
AUTHORS Omura.S., Ikeda.H., Ishikawa.J., Hanamoto.A., Takahashi.C., Shinose.M., Takahashi.Y., Horikawa.H., Nakazawa.H., Osonoe.T., Kikuchi.H., Shiba.T., Sakaki.Y. and Hattori.M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan  
(E-mail:bioinite.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun Ichikawa(\*2), Akihara Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi Osonoe(\*4), Yoshihiro Kishida(\*4), Hisashi Kikuchi(\*4), Tadayoshi Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7) and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermitilis.ls.kitasato-u.ac.jp.  
location/Qualifiers  
1. 320150  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227982"  
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinii"  
complement(134. 11953)  
/gene="olmA3"  
complement(134. 11953)  
/gene="olmA3"  
/note="SAV2897  
oligomycin gene cluster"  
/codon\_start=1  
/transl\_table=11  
/product="modular polyketide synthase"  
/protein\_id="BAC70608.1"  
/db\_xref="GI:29606547"  
/translation="MANEQDRLKRVSDLHRTORLRTIESQGTPIAIVGNACR YPGVSSPEDILMRVLTDETDAIGPPTDRGWDLDGLHPDHPGTCYAEAGGFVEDI ASFDAFGISPREAQSDMPQORLLTSEWLEHAGLDIHALRGSTGTVAGLNQD YGTLAAAPDGLSDYSGTNSVLSGRISYVGLGPAVTVDTACSSSLVTLHAAQ ALRSGCDLALAGATLTLSAVHVALSGRALAPDGRSKAFSAADGAGSGVGL AVERSDARRIGHRVLAIVLGSVAVQASNGLTAPNGPSQQRVIRQLANAGVSPAE VDVEARHTGSLGDPTEADALLATYQAKTAPRLMIGSLKSNIGHSGAAGVGVI KMQALRHGLPRTLHAEPTPNVDSGSAVELLETPARWPASDTPLAAVSSFGIS TNAHVLEQAPETPEAPEPTGEGATAPLLPVLVSARGSPALRAQARALEHL LARPGSADVALSLAVTRTGLEHRAVLAADRGLDLDALATDRPAEGLRSGPP VDGKVVFPQCGSQWAGMARELLDTSVFAAKAAACAATIEEFVDWKLVDLREKPG AASMERLEVQPVLFTVWVSLAEWLRSITGVFPDAVSSSQGEIAAAHVAGGLTQDAA RVICIRGALLAETLIVGKVASVALPDAVERLTRWADRLSVAGNPRIVAVAGSTV DALAEFTAEACARDNVRIVAATVPTHCALVDPLRELLDLAPVAPRTVPLYSTV TGRDLTAAMDAGVYNDTRAPLPEPVYRALLAEHSAFVSSAPHPVLAAMEGVEQTV AAGASGVVVSIESRDEGLGRTASLAKAHGGVYVDWVPVAFAGTAVTLPYAFQ RTRYWTAGDRTGDSVGLSDVHPLGALVRMADGDGAVLTGSLRAQTWLAGHG VSDRVIFPGTGFVELAVLAGQVCGGRIEELTLPLVVPESGAVVQAQVEADETG ARASVSVSRADDAADTWTWRHSGVLAEPGEPEGITGTWPPEGASVDSIGMYDR



Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, D.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with differing host specificities  
 Nature 417 (6887), 459-463 (2002)  
 22022145  
 12024217  
 2 (bases 1 to 12201)  
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Queglio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

Direct Submission  
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil

## FEATURES

Location/Qualifiers  
 1..12201  
 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 33913"  
 /db\_xref="ATCC:33913"  
 /db\_xref="taxon:190485"  
 /note="Pathovar: campestris"  
 68..1018  
 /gene="XCC1538"  
 68..1018  
 /gene="XCC1538"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA040833.1"  
 /db\_xref="GI:21112614"  
 translation="MRWOLRGAGTGVGVILVWVACSGTSGSGAAEPSSARADRG  
 GYAALAHARARAAQLHAPDGLSYTGSGVRVRCVRVADPHADIVLPAGPAQLGN  
 LITLPEGLQRQAAGTGVTLHGKPVTEVWLAERAGPGRLRIDAREFVYVQTGAL  
 YGWRPDPAALQAGELGYFFDPAHWRIDQWQPYDAPRATLITLSGTPLTVQVPGS  
 AVFTNRQYRLRPLAQGDARLFPVFADRTSGKESGGYARLYTADPRDGRVLDEN  
 LAENPPCALTPHVVCPTAPPDNRDLAVNAGEKTYRPSGR"  
 complement (1028..1603)  
 /gene="XCC1539"  
 complement (1028..1603)  
 /gene="XCC1539"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA040834.1"  
 /db\_xref="GI:21112615"

/translation="MSDAVLFPSPPTTAPLSYESRLARLTNRNIDMVVHCTELPDMA  
 MARYGERVLYDSGTSGNHYIYIDRNGSIQOYVALERVAHHVRGHNPHTLGIELVNSG  
 RYHMLDSRHOQREPYDPAQINALVTLLQLEQLPSVRRMAGHEALDTTQEAASDD  
 PSRKIQKRDPGGLFPWHERIETAVRWVHAPR"  
 1627..2218  
 /gene="XCC1540"  
 1627..2218  
 /gene="XCC1540"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA040835.1"  
 /db\_xref="GI:21112616"  
 translation="MALNPDSPLGKMATLPQTGQVTVGVVPARDVAMLEVDAAQAT  
 TGIGLVDRYKSGSGKRGITLLOAEHLPTVIAALAGQAALAPATIRRNLYVVGILIAL  
 KGRFRIGEVELEGTDPDCPSRMEALPGGYNAMRGHGLCARILKGGVIRIGDTV  
 QAL"  
 2115..2642  
 /gene="XCC1541"  
 2115..2642  
 /gene="XCC1541"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA040836.1"  
 /db\_xref="GI:21112617"  
 translation="MSRGHCILAHFGSGPDALKVTALEAVERLGMWTHERPDPFTDL  
 ARDLGQGVDRGLRLLEIARATKGPVVLASGSLGSYIAAQVSLQVPTRALFLM  
 VPPTKMGPLPALDAAAVPISIVAHWDELIPADVIWAQARSARLLLVDDGHRLLGAH  
 VQASRAFAELQSL"  
 2687..3193  
 /gene="XCC1542"  
 2687..3193  
 /gene="XCC1542"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="acetyltransferase"  
 /protein\_id="AA040837.1"  
 /db\_xref="GI:21112618"  
 translation="MSVLIREASDGAALIEVLTMVAFMRAEHSRHSQHVIAALRND  
 DALSLVADHDGYVWGHVLAIVSPVLSGSGWALGPLAVPGPHQGLGTLRLVQAA  
 LATURERGAAGCAVHSGPFRRLGFAVEPGLSVDPVASELQALAFGLRLPLADVA  
 YHPAFGLG"  
 3368..4327  
 /gene="XCC1543"  
 3368..4327  
 /gene="XCC1543"  
 /note="putative; ORF located using Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AA040838.1"  
 /db\_xref="GI:21112619"  
 translation="MQHDDQPLPQTAPHGRVISHLGETDQDATTLLASIKTRLRHND  
 RSAAMSHLLQLVDALWAMELGRFLNRNLGNWHTQVLYTQAGTLVQQTTPALLEYE  
 LFTPTVALTRFRFGIYRTQALRPGMTLASLPCGMADVLSLDYGCQCPDVRLLG  
 IDLPDQDALRLAQOQGLEQLTQREDAPGEPNSVDVLSNGNLNLYEPDDARY  
 IALYRPFDRLRPGGTIVASFLTPPPLCAESPWKLLDIDGAALVLQSQVFIDVVQAR  
 WASFRTHAQRTQLESAGTDFRFDDBRASIFPTVIARKPMSD"  
 4693..6306  
 /gene="XCC1544"  
 4693..6306  
 /gene="XCC1544"  
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11







This page Blank (usps)